# DOCUMENT RESUME

ED 183 583 TM 010 102

TITLE Curriculum Evaluation. Conference Papers.
INSTITUTION Australian Association for Research in Education, Melbourne.
PUB DATE Nov 77
NOTE 395p.; Papers presented at the Annual Conference of the Australian Association for Research in Education (Canberra, Australia, November 11-13, 1977); Some papers were not included; Omissions are noted in document

EDRS PRICE MF01/PC16 Plus Postage.
DESCRIPTORS Conference Reports; Course Evaluation; *Curriculum Evaluation; Educational Policy; *Educational Research; Educational Trends; *Evaluation Methods; Models; Program Evaluation; *Research Problems; Teacher Evaluation
IDENTIFIERS Australia

ABSTRACT

Over 40 papers presented at an Australian conference on curriculum evaluation are included in this publication. Among the topics are: evaluation models; the relationship of reliability to the amount of observation; Elementary Secondary Education Act Title I evaluation; designing policy-oriented research; methodology in evaluating teacher behavior; effects of school systems on socioeconomic attainments; trends in evaluation; evaluator selection; student satisfaction and school environment; use of criterion-referenced tests; school-based curricula; poly-ethnic environments; alternate methods of evaluating scores for college admission; decision making; ethnographic evaluation; preservice teacher education; evaluation problems at grade 12; implications of evaluation methods; mathematics textbooks; self-perception of student teachers; open education; pupil evaluation of student teachers; children's language usage; adolescent intellectual development; historical overview; limits of intervention programs for the mentally retarded; student transition to secondary schools; Aboriginal Australians; early childhood education; and anxiety and achievement. (MH)

***********************************************************************
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
***********************************************************************

ED183583

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

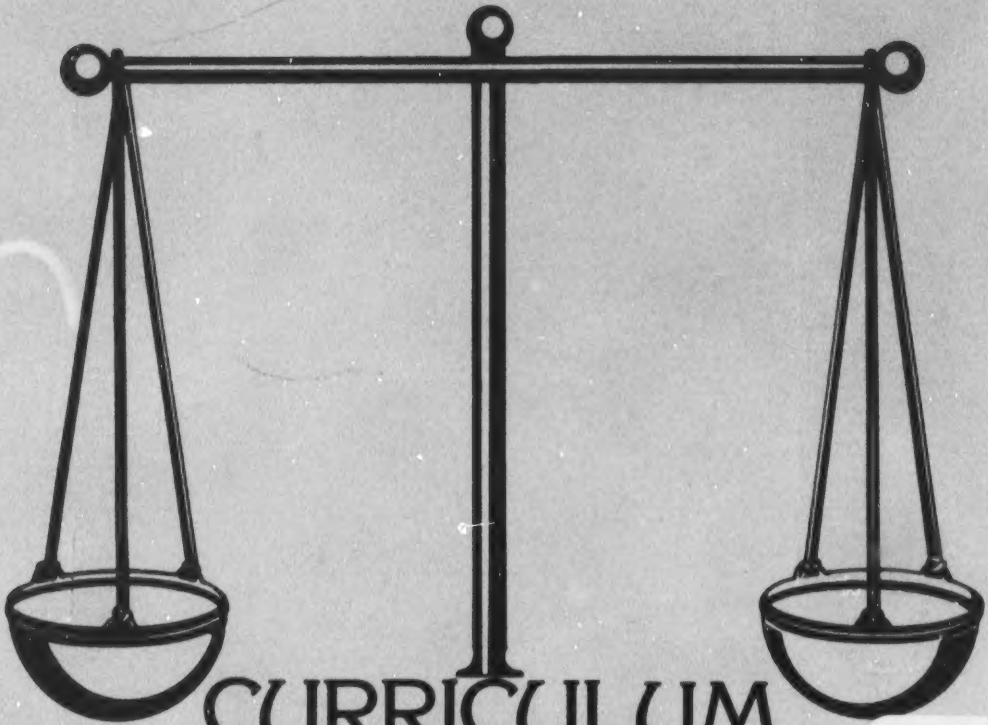
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# Australian Association For Research in Education

## 1977
## Annual Conference

## Canberra

DEC 1979
RECEIVED
ERIC/TM



# CURRICULUM EVALUATION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Phillip W. Hughes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM010 102

## Conference Papers

## 2

## FRIDAY 11 NOVEMBER

## 9.30 - 10.30am

| PAPERS | Rooms |
|---|---|
| G. Detterick (Gippsland I.A.E.)<br>Models of Schooling and Models of Evaluation | Seminar Room A |
| G. Rowley (La Trobe University)<br>The Relationship of Reliability in Classroom Research to the Amount of Observation | Ladies Drawing Room |
| R. Baldauf, Jr. (James Cook University)<br>A Systems Approach to Title I Evaluation: Implications for Small Cultural Groups | Seminar Room B |
| N. Baumgart, B. Low (Macquarie University)<br>Designing Policy Oriented Educational Research | Meeting Room |

NOTE: Some of the papers listed on these schedule sheets were missing before receipt of the document at this Clearinghouse. Omissions are noted.

# MODELS OF SCHOOLING AND MODELS OF EVALUATION

Dr Graham W. Dettrick
Gippsland Institute of Advanced
Education
Switchback Road
Churchill, Victoria 3842

Presented at the Annual Conference of the Australian Association for Research in Education, Canberra College of Advanced Education, November 11-13, 1977.

Less than two centuries of formal schooling in this country have left us with a legacy which is rich in educational values and the goals of schooling. Our philosophical and political heritage has given us the choice amongst Locke's ideal of an educated elite, schooling for the inculcation of habits of industry and piety, McArthur and Rusden's belief that schooling should "prevent vice and crime (Ely, 1977, pp.32-33)", Spencer's dream of schooling as preparation for complete living and Holden (1853, pp.1-6), who desired the existence of upward mobility. There have been others who have fancied "education for a skilled labour force (Ely 1977, pp.60-61)", equalization of opportunity (Karmel et al., 1973, Ch.3), schooling for national efficiency and international survival, and as an agent for transforming society. What our legacy lacks is any certain knowledge of which model of schooling best realizes any of these concepts of education. It is possible that it is time to choose a different model of schooling, or perhaps different models, but it is equally important to be aware of the crucial correspondence between a model of schooling and its model of evaluation. If the model of schooling is not matched with its model of evaluation, we are likely to become disillusioned. Disillusionment in the past has led to the somewhat fruitless search for various educational culprits. The most appropriate response is to see our model of schooling and our model of evaluation for what they are and to see how we need to change both models.

Our schools are almost entirely organized to give group instruction based on a uniform time schedule. This may be one efficient way to run a large organization, but its very convenience in dealing with large numbers of pupils leads inevitably to the measurement of individual achievement under conditions of standard exposure to instruction. Carroll (1963) has argued that where individual pupil achievement is the basic measure of schooling and where the school is committed to a uniform exposure to instruction, it is inevitable that aptitude and achievement will be highly correlated. Given a group of pupils with varying aptitude, and given the same exposure time to what is to be learned, it follows that their performance on an achievement test after instruction will vary on a distribution approaching that of a normal curve (Bloom, 1971).

It is crucial that the relationship between measurement of educational outcomes and the organization of school be kept in mind. No matter what ideal system of schooling is invented, if it relies upon standard time for instruction it seems inevitable that the outcome measure will be a distribution of individual pupil achievement along a curve that approaches the normal and which will be positively correlated with initial aptitude scores.

Our present model of schooling for the masses has come to us as a legacy from dame schools, charity schools, and "working schools" in the latter part of the eighteenth century. These humble beginnings provided the basis for the development of a mechanical system where subject matter and children were carefully graded and where there was little or no opportunity for the asking of questions, nor, of course, for the development of individuality. Recitation was the most common teaching practice. It is somewhat chilling to note that, on the basis of current investigations into teaching, Dreeben (1973) reports that recitation remains the predominant schooling practice, viz.,

> "Highly suggestive but by no means definitive evidence that runs back to the end of the nineteenth century indicates that teachers run classrooms; they do most of the talking and initiate most of the action. In the relatively short spans of time left to them, pupils mostly reply to questions and less frequently ask them. Evidence on the spatial distribution of talk locates most of it toward the front and center of classrooms (in both secondary schools and college classes). When the teacher talks, the chances are substantial that a large number of pupils will be addressed (not a small group), that the pupil who talks next will be sitting near the front or along a center strip, and that the teacher will be the following speaker (Adams & Biddle, 1970, pp.63-68; Sommer, 1969, p.118). Classrooms, in short, are run on the principle of the recitation (Hoetker & Ahlbrand, 1969, p.146).
>
> If systematic evidence about the past and about current practice, taking full account both of modal patterns and of variations, supports this generalization, we will have discovered a most remarkable phenomenon: the predominance of the recitation. Presumably there are many ways to teach a classroom full of children or to teach them in settings different in structure from classrooms, but for reasons as yet unknown these alternatives have not gained currency (p.465)."

The graded school was organized on the basis of years or quarter years of work. Pupils progressed or did not progress on the basis of frequent examinations. This graded system of mass production in education fitted in well with the economic and industrial ideas of the nineteenth century. Given too few teachers and educational facilities, it is not surprising that the graded school remained popular. This model, which we shall term the graded or classification model of schooling, was subtly able to restrict, select, and eliminate and, where necessary, "guide" through regular examinations. For the first fifty years of the twentieth century, this model has operated to limit opportunities to secondary academic schooling which represented a necessary prerequisite for entry to certain "white collar" occupations, the Public Service, or the university. In much the same manner, crowding in Paris schools and the need for selection gave impetus to Binet and Simon to develop the IQ test, a measure which could predict or select those pupils most likely to succeed in the existing school system; a job it has done well.

While there may have been leaps in scientific and technological development which have been accompanied by dramatic changes in social values, the classification model of schooling has persisted as McLuhan & Fiore (1967) have observed,

> "Today's television child is attuned to war, taxes, crime, bathing beauties - and is bewildered when he enters the nineteenth century environment that still characterizes the educational establishment where information is scarce but ordered and structured by fragmented, classified patterns, subjects, and schedules. It is naturally an environment much like any factory set-up with its inventories and assembly lines (p.18)."

Assessments of the classificational model of schooling seem to indicate that while it may act well as a screening system, it has a very limited effectiveness on equalizing educational opportunity. A major finding of the Coleman Report (Coleman, et al, 1966) was that there was a "relatively small amount of school-to-school variation that is not accounted for by differences in family background, indicating the small independent effect of variations in school facilities, curriculum, and staff upon achievements (p.235)." This major finding of Coleman could be seen as an enormous threat to the classification educational system model if we were prepared to take more than passing notice and were really concerned about equalizing opportunity because the implication of the Report is that our present major model of schooling in itself, does little other than permit, or possibly enhance the pupil's own initial ability to achieve. Corresponding conclusions are reached in the International Study of Achievement in Mathematics (Husén, 1967, pp.253-256), the Karmel Report (Karmel, et al., 1973, pp.16-20) and a review by Dreeben (1973, pp.455-456).

These findings should be regarded as neither a tragedy nor a surprise, since a strong argument may be made that this is what the classification model of schooling is all about. It may even be argued that, the better the school, the more pupils will vary in their individual achievement, which will be closely related to their initial aptitude. Good schooling, according to this model, should increase individual differences, not decrease them.

However, this is not the popular interpretation which has been placed on these findings. Rather it is often seen as an indictment of our schools. The result has been a hunt for the culprit. For some, the financing of schools is the culprit. Some have selected the resistance of the school system to innovation (De Lacey, 1974). Others select our techniques of educational evaluation, or decision-making, or centralized administration (Maclaine & Selby Smith, 1971, pp.6-12). Still others select educational research (Miller & Gordon, 1974). All agree something must be wrong. The view put forward here is that nothing is wrong except our reluctance to analyze the model of schooling and its concomitant model of evaluation.

Sieber (1968) led an attack on the U.S. school system saying it is "a vulnerable formalized organization with diffuse goals whose functionaries are quasi-professionals, and which is devoted to processing people within its boundaries." Sieber also indicated that while schools are obliged to deal with a larger and larger percentage of the U.S. school age population, rising from 68% of 16 and 17 year olds in 1947 to 88% in 1964, the schools have been unable to change their role to adjust to new demands arising from increasingly diversified backgrounds of children remaining at school. Sieber has blamed the lack of adjustment upon bureaucratization and decentralization. Not only have corresponding trends of increased enrolment occurred in Australian secondary schools, for example, the percentage of 15 to 19 year olds pursuing secondary education has risen from 11% in 1947 to 21% in 1959, 26% in 1966, and 36% in 1974, but corresponding culprits have surfaced in Australia. The quality of our teachers has been questioned (Miller, 1957; Warry & Fitzgerald, 1969; Karmel, et al., 1973, pp.111-112), the manner in which schools exist to process children has been critized (Connell, 1970), and by way of contrast with Sieber, the centralization of the Australian educational bureaucracies (Maclaine & Selby Smith, 1971, p.7) has come under attack.

Our classification model of schooling has an organization whose goal is to control the flow of students through the system, that is, it is the schools' job to select some students to move up the schooling ladder and to terminate, slow down, or change the direction to sideways for others. The reality of this model of schooling becomes apparent if we recall the development of special educational classes as one educational innovation which swept this country in the late fifties and sixties. To the regular academic/technical streams we added "special" or "modified" courses such as "Allied Mathematics", "Applied Science", and so on. The reason for this appears to be that this mechanism removes from the regular classroom groups of children who are seen as incompatible with their cohorts. During the seventies demand and supply of remedial education has increased. The effect of this is also to remove children who seem academically incompatible with their cohorts. We might speculate then, that the classification schooling model will adopt those innovations which increase the schools' ability to operate as a regulatory system handling a flow of students. In turn, schools will be reluctant to adopt those innovations which are anti-regulatory - such as truly individualized instruction - in fact any innovation which would treat individual students as individuals rather than as units within a flow system.

Here we have a basic conflict. On one hand we have a classification model of schooling organized to process groups of students at differential rates and using, quite properly, evaluation techniques to discriminate groups by differential learning rates. This model is now faced with not just more pupils, but an increasingly universal population target rather than a selected one. This increase in school population may have an interesting effect. As the percentage of the school-age population remaining at school rises, the mean aptitude of pupils must be lowered because lower aptitude children will form a proportionately larger part of the increased pupil intake. "Standards" of schools will appear to decline and a public outcry may be anticipated from the increasingly educated clientele amongst pupils and parents as the schools appear to be failing.

One would speculate that the immediate response would be the restoration of the classification steady state by processes assisting greater selectivity, viz., public, independent and select state schools will become more popular and gain better financial support, special school curricula will be devised, more remedial facilities will become available, and more finance may be allocated for materials and support services.

The tendency for schooling to have to accommodate to an increasingly universal population developed earlier in the United States than in Britain or Australia and it is in the United States where we have the first clear signs of a massive witch-hunt. For some time prior to the Sputnik era, schools had been dealing with the lowering average aptitude by offering a wider range of terminal options which were more general, "easier", and more vocationally oriented than the "academic" courses of study which were still maintained. Following the shock of failure in the space race on October 4, 1957, a search for a culprit began. Since the persons doing the hunting were academics, principally scientists and mathematicians, it is not surprising that the culprit was determined as the substance and method of the disciplines taught in primary and secondary schools (Bruner, 1963).

For the millions of dollars spent to overcome the chimera, the curriculum developments which occurred, for example CHEM study, CBA, B.S.C.S., E.S.C.P. and P.S.S.C. did not gain widespread diffusion and the percentage of pupils benefitting from the effort was comparatively small. However, this should not be surprising since many of the curricula, even those designed for primary and junior high school, were written for pupils who were college bound. These curricula tended to reinforce the selectivity of the classification model. Not only was the percentage of any high school population studying these courses small, but those familiar with the achivement tests produced will recall that the tests are norm-referenced. In Britain, where curriculum development thrusts occurred later than in the U.S., curriculum projects not only devoted substantial attention to the substance and method of particular disciplines, but effort was also directed to alter the traditional relationship between the teacher and the knowledge base. Delamont (1976, p.47) cites the case of the "Nuffield" approach to science teaching which emphasized individualized discovery-based learning. This curriculum, which was hailed as revolutionary and which was thought to be sweeping the schools, was found to be taught in the "true Nuffield manner" by only 17 percent of the teachers sampled (Galton, 1976).

In Australia, criticism of the substance of the curriculum has also occurred. Responses have been either to adopt curriculum packages, e.g. CHEM Study, P.S.S.C., modify curriculum packages, e.g. Web of Life, V.U.S.E.B. Chemistry course, or to produce courses indigenously such as the Butler-Messel science programme in N.S.W. The implementation of this latter programme was accompanied by a restructuring of the secondary curriculum in N.S.W. - the Wyndham Scheme (Wyndham et al., 1957) - which manifestly supported a classification model of schooling. High school curriculum development in Queensland in the late sixties also showed a structural as well as a substantive support for the classification model, for example in science junior high school "Special" Science or Science A or Science A and Science B was followed in senior high school by one of more of the U.S. packaged programmes. The dominant instructional methodology used with any of the newer secondary curricula is the traditional class teaching approach which is also a notable feature of our primary schools (Bassett, 1964, p.23). Minor variations of this style, such as intra-class groupings, do not markedly alter the most common feature of the classification model which is that teachers appear to be instructing pupils for a common level of attainment. This, of course, is consistent with the classification model.

Curriculum developments which do not easily fit the classification model, e.g. J.S.S.P. (Junior Secondary Science Project) and A.S.E.P. (Australian Science Education Project) have not proved to be popular because of the increased stress upon the attainment or mastery of specified objectives, together with a more individualized approach where pupils may take different periods of time to master equivalent material. Many behaviour oriented learning psychologists (Bloom 1968; Carroll, 1963) have been among those arguing for such an approach through more precise definition of models and of educational goals in behavioural terms. Mastery of any given subject matter or skill can be specified in fairly precise behavioural terms and is therefore more subject to measurement. This approach might lead to schools where each pupil is expected to master or reach specified degrees of proficiency with subject matter. This, in turn, would change drastically the nature of the school. If the outcome measure of instruction were to be mastery, then the type of instruction and the time for it would have to be varied according to individual pupil characteristics. The outcome then would be, hopefully, that a majority of students would achieve a specified level of mastery and instead of aptitude and achievement being highly correlated as a result of uniform instruction, there would be little or no correlation between initial aptitude and mastery measures, but time to mastery would still correlate with aptitude scores. In this instance the model of evaluation would totally change the model of schooling.

Some have selected educational evaluation as a culprit and charge almost any educational test with not adequately measuring pupils' performances because, it is argued, pupils "real" performance potential is higher than that measured. Norm-referenced tests at the state level have fallen into severe disfavour, and the norm-referenced disfavour symptom has become quite obvious in schools. However, if one wanted to select the fast from the slow learner, then it makes a good deal of sense to build tests whose items would make just that discrimination. The school and state achievement tests which have been used for decades do this job very well even though some of the tests may have more or less aesthetic appeal to people who consider themselves relatively expert in educational measurement. While no adequate popular substitute for norm-referenced tests has been developed the criticism goes on. It should be remembered that a change in the evaluation process such as the elimination of matriculation and other terminal examinations does not create a different educational model, but merely alters the system of reckoning. Measurement can not invent, it only accounts.

The real issue is whether or not what should be measured in the classification model of schooling is what we now want to measure. Glaser(1968a, 1968b), Glaser and Cox (1968) and Lindvall and Cox (1968) have argued for the importance of tests which measure how well the individual pupil has attained some criterion on what they call "criterion-referenced" measures. These measures would depend upon an absolute standard of learning and would give explicit information about what individuals can or cannot do, subsequent to some form of instruction. Tyler (1967) has argued that we need to go even further and design measurement programmes to assess what "everyone", in the statistical sense, can do. It seems important not only to know the quality of the nett educational product in the sense that we know what specific educational skills are at pupils' command, say at tenth grade level, but also that we know what our human resources are, what we are obtaining for the millions spent in the name of education, and where we need to change our thinking or modify our actions. Those familiar with the National Assessment of Educational Progress will know that in the United States significant developments in this direction have already occurred on an amount which is a small fraction of the annual education budget (U.S. $6 million in 1973 and U.S. $4.5 million in 1974 for a sample size of approximately 90,000 persons). What the foregoing suggests is that our classification model of schooling which has acted as a complex set of filters to organize a pyramid of pupil attainment within limited educational opportunities may no longer be the appropriate model. This view should not be seen as novel because corresponding suggestions arose at an A.C.E.R. conference in 1962 (Bassett, 1964) but the impact of these ideas has been slight.

Another culprit in the evaluation field has been that of curriculum evaluation. Some of the borrowed or modified curricula we have been using for some years have become progressively less popular in schools. This unpopularity is not just a matter of the curriculum being old fashioned or out of date, but that the traditional uniform instructional model has tended to produce declining results. As the percentage of school age population has increased and the task of the school is increasingly being recognized as that of reaching all children including those who have not been making appreciable progress in learning before, more pupils are being conveyed through the various sorting filters and, since it is becoming increasingly difficult to fit pupils into curriculum courses, there is a growing tendency to select courses to fit categories of students - a further mechanism to maintain relative homeostatis for the classification model. There has been some feeling (Gagné, 1967) that we lack an adequate base for making any other determination of the curriculum except on the basis of what pupils have learned and at what level of mastery. In order to demonstrate mastery empirically, one must consider how this can be measured. This leads inevitably to a discussion of goals in measurable terms, or, to use current terminology, behavioural goals. We have floundered in curriculum evaluation on some very difficult methodological problems such as the inability in practice to assign students or curricula randomly, to change the sequence of teaching for purposes of experimental design, or to locate comparable control groups. The methodological problems are difficult, for students in experimental curricula do better on certain things and less well on some conventional matters (Herron, 1966). It is almost impossible to answer the question of which curriculum, the old or the new, is better because determining what is better calls for a criterion referenced value judgment mixed with data. While we maintain the current schooling model, it is doubtful whether varying the sequence and content of curricula will significantly alter student achievement.

Some critics feel that our educational research efforts have been too pure, too impure, too applied, not applied enough, not sufficiently on target, or too organized. In any event, they claim research efforts have not produced results which have had much effect upon the schooling process (Miller & Gordon, 1974; Brinkworth, 1977; Travers, 1973). An encyclopaedic review of research such as that in the Second Handbook of Research on Teaching (Travers, 1973) provides some clues to explain the lack of impact. Not only is much of the research little known because it is reaching the "workface" through the process of diffusion rather than actual dissemination but, much of the research evidence appears to be essentially negative (Schutz, 1969; Dunkin & Biddle, 1974, pp.11-30) or exposes conflicts arising from the clash of various paradigms (Nuthall & Snook, 1973). It is not surprising then, that the education system is reluctant to adopt innovations. Until there is strong evidence favouring a particular approach, the classification system is very likely to go on doing what it has always done, namely to adopt those changes which make its present job easier to accomplish. Guba (1968) and Stufflebeam (1968) call for new educational research designs for they believe that specifications such as those given by Campbell and Stanley (1963) in their paper on experimental designs are often unrealistic. Obtaining the cooperation of a teacher, school, or school system for a research venture is often a non-random unrepresentative sampling problem. Random selection of pupils or classrooms is almost impossible to achieve and even where it is done, functionaries ordered to carry out a treatment may invalidate the data by appearing to cooperate while they proceed to do exactly as they have always done. This ultimate lack of control may help to account for many non-significant findings in otherwise well designed and executed investigations. Utilization of less complicated designs to pursue one variable in a large enough sample of a population may resist those effects which operate in real school life.

For example, the National Assessment of Educational Progress in the U.S. is gradually accumulating information which is developing a clearer picture of what the school age population has mastered at ages 9, 13 and 17 and what remains with young adults (age between 26 and 35). Another alternative is to pursue a study in the school situation where the investigator may stay continuously in contact with all phases of the investigation, consequently reducing the size of the sample, but putting maximum effort into control, observation and recording. This is an expensive design which would usually require more than one investigator to control treatments for as few as thirty pupils. Even then, interactive effects of experimenter, experiment, observers, pupils and school staff may make the sample representative of itself only as a result of the research project. A third alternative is to experiment with one student, one class, one school or one school system, to build a working model of what one wants to happen in education and then see how well the model works in terms of clearly specified criteria.

The dissatisfaction with the unrealistic demands of many experimental designs has caused some to argue for a return to field studies or more naturalistic studies (Belanger, 1969). Educational research has leaned heavily upon psychological research design, which has been derived from agricultural research and uniform treatment; conditions which scarcely apply to investigations with pupils in the complex interacting conditions of the school room. In our research we may also have overlooked the point that many so-called experiments in curriculum may not be experimental research at all. Given a classification model of schooling which selects and processes cohorts of pupils according to their learning rate and given approximately equal instructional exposure, then the outcome is fixed on any sort of norm-referenced achievement test. Most of the variance can be accounted for by method of selection, type of exposure, learning rate criterion, and the model of evaluation. Given the relationship between the classification model of schooling and its model of evaluation the effect of other variables is relatively small. Until the effect of the relationship between the classification model and its corresponding norm-based model of evaluation is fully realized, many experimental efforts will be pseudo-scientific in the sense that the larger order educational variables are already controlled by the model itself. Until a methodology to deal with the situation is invented, we might advance our knowledge by moving away from experimental designs to more value specific, criterion referenced evaluation models in schooling systems which pay greater attention to a mastery model of schooling.

Political, economic, and professional criticism have branded the structural and decision making components of our education systems as culprits responsible for rigidity, resistance to innovation, paternalism, conformity, and a host of other ills. Considerable attention has been paid to the operation of these systems and their components and the manner in which it is believed they may be refurbished to function more effectively (Jecks, 1973). However, comparatively little attention has been devoted to the detailed specification and attainment of those goals which these systems could help us reach. A thorough use of systems analysis would cause educational administrators to face, more acutely than before, what it is that is expected of the schooling system, and how much the various performance goals are valued. These values may be translated into people, finance, and measures of pupils' progress towards a specified goal. In a criterion-referenced model, one cannot avoid specifying purposes, placing values on them and measuring goal attainment. Our history of educational decision-making is not rich in examples of the clear statement of educational goals, their systematic implementation and their measurement upon the basis of some agreed criteria. If one was to take the Wyndham Report, for example, it concerns itself principally with renaming courses, adding to the number of core studies subject fields, and suggesting different ways of shuffling pupils around (Wyndham, et al., 1957, pp.63-98). Systematic plans for the implementation of vague educational goals (pp.56-62) and the measurement of the level of attainment of these goals is not mentioned. This common "innovative" procedure may be anticipated while we maintain a classification model of schooling which concerns itself almost entirely with how pupils rank and not what pupils can do at what level of competence.

It is worth emphasizing that it is a much simpler matter to weigh alternatives when the goal is specified clearly than it is to weigh alternatives for an education system whose goals are viewed quite differently by segments of the population when none of the goals are expressed in precise terms. However, while there are benefits, there may be pitfalls. Among the anticipated problems Hartley (1969) warns against goal distortion by which he means that organizations using methods like systems analysis and criterion referenced evaluation may place greater emphasis on goals which are easier to measure to the detriment of those goals which are difficult to assess. Examples of strategies which have been implemented to circumvent these possibilities may be found in ongoing planning for the U.S. National Assessment of Educational Progress (NAEP, 1973b). Other educators warn of problems which may arise if administrators improperly interpret and subvert the use of results (NAEP, 1973a, p.4; NAEP, 1973c, p.5).

Educational decision-making stresses economic data (Karmel, et al., 1973) and has considered the economic benefit in terms of a "national investment in human capital (Martin, et al., 1964, p.4)" almost computing rates of return for different degrees of education. Until the outcome measures are clarified in such economic analyses, the same model of schooling may continue. For example, if relative academic achievement is used as a measure of output, the norm-referenced classification model remains. On the other hand it is possible to adopt a different model and ask a criterion based question, for example, what would it cost to bring X percent of the population to this level of mastery. This type of decision-making is beginning to flow from the National Assessment of Educational Progress in the U.S. (N.A.E.P. Special Feature Jan.-Feb. 1974). Terrill H. Bell, U.S. Commissioner for Education (NAEP, 1975) has suggested that as a result of a management-by-objectives approach supplied by data from the NAEP "education managers can identify needs, set priorities, and establish a clear sense of direction (p.5)" and states could deal with differences which have been "shouting for action for 30 or 40 years (p.5)." Many states are moving ahead rapidly to tackle the issues and complexity of "minimal competency". Already four states: Florida, California, Colorado, and Virginia, have passed the necessary legislation (NAEP, 1976, p.3). From NAEP surveys taken in the last seven years in reading, writing, science, and mathematics, a "back to the basics" approach in the U.S. may be made on a selective basis (NAEP 1976, p.6-7) and effort and finance directed where it will produce the most beneficial results. No such approach is currently possible in Australia.

The Karmel Report (Karmel, et al., 1973, p.11, pp.16-24) places considerable emphasis on equality of educational opportunity in that pupils should be free of restriction to the circumstances of a pupils' family and have equal access to schools of equal standards. The success of any attempts, no matter what they are, by "some altering of the balance of expenditure in earlier stages to consolidate a more equal basic achievement (p.11)" will not be realized while we are currently committed to a model of education based on pupil sorting following equal exposure to instruction. Of equal concern is that while we persist with our present models of schooling and of evaluation we have no adequate mechanism for measuring the success or failure of plans implemented to realize the hopes of the Report.

A very persuasive argument for a different model of evaluation, one of mastery of learning, has been put forward by Bloom (1968), Carroll (1963), Glaser (1968) and Glaser and Nitko (1971) which if implemented would mean a new model of evaluation based upon mastery in specified units. The mastery model requires specific educational objectives, specified various routes and sequences to mastery, and specified measures of competency. In time, a mastery model of evaluation should permit far more individualization of the routes to mastery based upon learner characteristics and interests. The curriculum would have to undergo a logical analysis and evaluation which it has sadly lacked. An accompanying model of schooling would not be committed to the concept of a sequential system of schooling by grade, year or form. The model would not be able to operate on the rationale that every pupil should "go as far as his abilities warrant" or that he be provided with the appropriate curriculum once his prospective future role is determined. There would be no "weeding out" process and no channeling into different courses such as academic, vocational courses, or business courses, nor would there be one form of mathematics for "business students" and another for "academic students". The mastery model of schooling requires clear and definite specification of educational goals so that these may readily form the basis for criterion referenced measurement. A vague statement of goals leaves little information about what the teacher and pupil are to look for, or provide the necessary guidance to determine the subject matter structure and component tasks in the system which education managers must devise to reach the specified goals. It is necessary to remind ourselves that measurement must be tied to the educational goals and that much of the measurement will be invented in response to the setting of goals. If, for example we determine that all persons should reach certain performance levels in relation to certain goals and neither equal exposure time to instruction nor grade, year, or form sequencing were moderating factors, we could have pupils of a wide range of ages and backgrounds taking a unit of learning together providing that attainment of specified goals permits a variety of instructional treatments at differential rates as individual differences of students are taken into account.

The non-graded school is an attempt to move in this direction but this will have limited effectiveness while the classification model of schooling remains. Other trends such as that to "open education", continuous assessment, and the disappearance of state-wide selection examinations in favour of some form of moderated school-based evaluation, are likely to make little discernible difference since these merely change minor conditions operating while the basic models of schooling and of evaluation are retained.

It seems unrealistic to persist with the classification model into the 1980s. We are faced with an altering reality, that of extended schooling for all and where education of one form or another will increasingly become an integral component of leisure and working life. The time when the process of education could be treated as a one-time experience of our youth is almost past (Duke, 1976; Faure, et al., 1972; Kangan, et al., 1974). In moving from the classification model as the predominant model to more pluralistic changing models, we must carefully avoid the intermediate error of eliminating some of the characteristics of the classification model to become trapped in a pathless unassessable rhetoric of estimable intentions, pious hopes, good feelings, and appealing intuitions. Models of education must provide clear goals and thoughtfully matched evaluative criteria. Where shortcomings in measurement of goal attainment exist, new processes of evaluation will need to be invented. With different models of schooling we will need to evaluate how well each succeeds in reaching the specified goals for its target population; and at what cost: educational, social and economic, in terms of the values and perspectives which arise from each model. I have no doubt that some form of national assessment programme would assist such developments.

Adams, R.S., & Biddle, B.J. Realities of teaching. New York, Holt Rinehart & Winston, 1970.

American Institute of Biological Sciences Biological Sciences Curriculum Study (B.S.C.S.) High School Biology (Green Version, Yellow Version & Blue Version). Chicago, Rand McNally & Co., 1963.

Bassett, G.W. Each one is different. Hawthorn, Vic., A.C.E.R., 1964.

Belanger, M. Learning studies in science education. Review of Educational Research, 1969, 39, 377-395.

Bloom, B.S. Learning for mastery. Evaluation comment. Los Angeles: U.C.L.A. Center for the Study of Evaluation of Instructional Programs, 1968. 1, No.2.

Bloom, B.S. Individual differences in school achievement: A vanishing point? (Phi Delta Kappa Monograph Series), Bloomington, Ind., Phi Delta Kappa Inc., 1971.

Brinkworth, P.C. The conduct and dissemination of research in mathematics education in Australia. In Research in Mathematics Education in Australia 1977 Volume 1. Melbourne, Vic., Mathematics Education Research Group of Australia, 1977.

Bruner, J.S. The process of education. New York. Random House, Inc., 1960.

Campbell, D.T. & Stanley, J. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.) Handbook of Research on Teaching. Chicago, Rand McNally, 1963, 171-246.

Carroll, J. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Coleman, J.S. Equality of educational opportunity. Washington, D.C., National Center for Educational Statistics. U.S. Govt. Printing Office, 1966.

Connell, W.F. Myths and traditions in Australian education. The Australian Journal of Education, 1970, 14(3), 253-264.

de Lacey, P. So many lessons to learn: Failure in Australian education. Harmondsworth, Mddx., Penguin Books, 1974.

Delamont, S. Interaction in the classroom. Lond., Methuen & Co. Ltd., 1976.

Dreeben, R. The school as a workplace. In R.M.W. Travers (Ed.) Second Handbook of Research on Teaching. Chicago, Rand McNally & Co., 1973, 450-473.

Duke, C. Life-long education: an Australian prognosis. (Occasional Papers in Continuing Education, No.13). Canberra, Centre for Continuing Education, A.N.U., 1976.

Dunkin, M.J. & Biddle, B.J. The study of teaching. New York, Holt, Rinehart & Winston Inc., 1974.

Ely, M.J. The rhetorical and real: Continuity of Australian educational history. Unpublished monograph, Gippsland Institute of Advanced Education, 1977. (Expected publication: Sydney, Alternative Press, 1978).

Faure, E., et al. Learning to be: The world of education today and tomorrow. Paris, UNESCO, 1972.

Galton, M.J. et al. Process and products of science teaching. Lond., Macmillan, 1976.

Gagné, R.M. Curriculum research and the promotion of learning, in R.W. Tyler, R.M. Gagné, & M. Scriven (Eds.) Perspectives of Curriculum Evaluation. Chicago, Rand McNally, 1967, 19-38.

Glaser, R. Adapting the elementary school curriculum to individual performance. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J. Educational Testing Service, 1968(a).

Glaser, R. Evaluation of instruction and changing educational models. Reprint 46, Learning R. & D. Center, University of Pittsburgh, 1968(b).

Glaser, R. & Cox, R.D. Criteria-referenced testing for the measurement of educational outcomes. In R. Weisgerber (Ed.) Instructional Process and Media Innovation. Chicago, Rand McNally, 1968, 545-550.

Glaser, R. & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.) Educational Measurement (2nd ed.), Washington, D.C., American Council on Education, 1971.

Guba, E.G. Development, diffusion and evaluation. In T.L. Eidell and J.M. Kitchel (Eds.) Knowledge Production and Utilization in Educational Administration. Columbus, Ohio, University Council for Educational Administration, 1968.

Harris, M.F. et al. Investigating the earth (E.S.C.P.) Rev.ed. Boston, Houghton Mifflin Co., 1973.

Hartley, H.J. Limitations of systems analysis. Phi Delta Kappan. May 1969, 515-519.

Hellyer, A.M. The Wyndham report. Unpublished Master of Education thesis, University of Sydney, 1968.

Herron, H.J. Evaluation of new curricula. Journal of Research in Science Teaching, 1966, 4, 159-170.

Hill, J.E. How schools can apply systems analysis. Bloomington, Ind., Phi Delta Kappa Educational Foundation, 1972.

Hoetken, J. & Ahlbrand, W.P.J. The persistence of the recitation. American Educational Journal, 1969, 6(2), 145-167.

Holden, G.R. Labour and capital. Sydney, 1853.

Husén, T. (Ed.) International study of achievement in mathematics: A comparison of twelve countries. (Volume 1). New York, Wiley & Sons, 1967.

Husén T. The learning society. London, Methuen & Co. Ltd., 1974.

Jecks, D.A. Major aspects of the organization and administration of Australian education. In A.G. Maclaine & R. Selby Smith (Eds.) Fundamental Issues in Australian Education. Sydney, Ian Novak, 1971, 134-175.

Kangan, M., et al. TAFE in Australia: Report on needs in technical and further education Volume 1: Report. Canberra, Aust. Govt. Publishing Service, 1974.

Karmel, P., et al. Schools in Australia: Report of the Interim Committee for the Australian Schools Commission (Karmel Report). Canberra, Aust. Govt. Publishing Service, 1973.

Lindvall, C.M. & Cox, R.C. Evaluation as a tool in curriculum development: the IPI evaluation program. Learning R. & D. Center, University of Pittsburgh, 1968.

Maclaine, A.G. The educational challenge of change. In A.G. Maclaine & R. Selby Smith (Eds.) Fundamental Issues in Australian Education. Sydney, Ian Novak, 1971.

Maclaine, A.G. & Selby Smith, R. (Eds.) Fundamental Issues in Australian Education. Sydney, Ian Novak, 1971.

McLuhan, M. & Fiore, Q. The medium is the massage. Harmondsworth, Mddx., Penguin Books, 1967.

Martin, L.H., et al. Tertiary Education in Australia: Report of the Committee on the Future of Tertiary Education in Australia to the Australian Universities Commission Volume 1. (The Martin Report). Canberra, Commonwealth Govt. Printer, 1964.

Miller, La Mar P., & Gordon E.W. Equality of educational opportunity. New York, A.M.S. Press Inc., 1974.

Miller, T.W.G. Changes in secondary education in Australia. International Review of Education, 1967, 13(3), p.272.

National Assessment of Educational Progress. NAEP Newsletter, 1973(a), 6(5), 4.

National Assessment of Educational Progress. NAEP Newsletter, 1973(b), 6(7), 4.

National Assessment of Educational Progress. NAEP Newsletter, 1973(c), 6(8), 5.

National Assessment of Educational Progress. NAEP Newsletter, 1974, 7(4), 5.

National Assessment of Educational Progress. NAEP Special Feature, Jan-Feb., 1975.

National Assessment of Educational Progress. NAEP Newsletter, 1976, 9(4).

Nuthall, G. & Snook, I. Contemporary models of teaching. In R.M.W. Travers (Ed.) Second Handbook of Research in Teaching. Chicago, Rand McNally, 1973, pp.47-76.

Physical Science Study Committee (P.S.S.C.) Physics 2nd ed. Boston: D.C. Heath & Co., 1965.

Pemental, G.C. (Ed.) Chemistry: An experimental science. (CHEM study). San Francisco: W.H. Freeman & Co., 1960.

Schultz, R.E. Methodological issues in curriculum research. Review of Educational Research 1969, 39, 359-366.

Sieber, S.D. Organizational influences on innovative roles. In T.L. Eidell & J.M. Kitchell (Eds.) Knowledge Production and Utilization in Educational Administration. Columbus, Ohio, University Council for Educational Administration, 1968, 121-142.

Sommer, R. Personal space. Englewood Cliffs, N.J. Prentice Hall, 1969.

Stranks, D.R., et al. Chemistry: A structural view, 2nd ed. (V.U.S.E.B. Chemistry Course). Melb., Melb. University Press, 1970.

Stufflebeam, D.L. Evaluation as enlightenment for decision making. Address delivered at Working Conference on Assessment Theory. Sarasota, Florida, Jan.19, 1968. Columbus, Ohio. Ohio State University Evaluation Center, 1968.

Travers, R.M.W. (Ed.) Second handbook of research on teaching. (American Educational Research Association Project). Chicago, Rand McNally, 1973.

Tyler, R.W. The objectives and plans for a national assessment of educational progress. Journal of Educational Measurement, 1966, 3, 1-4.

Tyler, R.W. Changing concepts of educational evaluation. In Perspectives of curriculum evaluation. (A.E.R.A. Monograph Series on Curriculum Evaluation, No.1). Chicago, Rand McNally, 1967.

Warry, R.W. & Fitzgerald, R.T. The new R's in the Australian primary curriculum. Quarterly Review of Australian Education, 1969, 2(4), p.18.

Wyndham, H.S., et al. Report of the committee to survey secondary education in New South Wales. (Wyndham Report), Sydney, Govt. Printer, 1958.

# THE RELATIONSHIP OF RELIABILITY IN CLASSROOM RESEARCH TO THE AMOUNT OF OBSERVATION: AN EXTENSION OF THE SPEARMAN-BROWN FORMULA.

Glenn Rowley
La Trobe University

## ABSTRACT

The reliabilities of various observational measures were determined, and the influence of both the number and the length of the observation periods on reliability was examined, both separately and jointly. It was demonstrated, as expected, that the reliability increases as either the number or the length of the observation periods is increased, and that, if total observation time is held constant, the reliability is greater for a larger number of shorter observation periods. Based on these findings, some practical recommendations were made concerning the scheduling of classroom observation. The relationship between reliability and number of observation periods may be accounted for by the well-known Spearman-Brown formula, but this relationship was shown to be unable to account for the relationship between reliability and the length of observation periods — as the length is increased, the reliability increases considerably less quickly than uncritical use of the Spearman-Brown formula would lead one to expect. Based on a single simplifying assumption, an explanation was proposed which accounts simply and accurately for the observed relationship, and leads to a variant of the Spearman-Brown formula which may have wider application.

Direct observation of teachers teaching, and of learners learning, have come, in recent years, to be regarded as important elements of research aimed at improving our understanding of the teaching-learning process. Of this research, a large part has been essentially quantitative, as evidenced by a quick glance through a review such as that by Rosenshine (1971), in which 51 essentially correlational studies are reported. Aside from the generally inconclusive nature of most of the research reviewed [1] one's attention is drawn immediately to the lack of impact which measurement theory has had upon the practice of classroom observation. For it is clear that a good deal of measurement goes on when teachers and pupils are observed, whether it be in counting the frequency of occurrence of particular behaviors, measuring the time devoted to particular activities, or computing the now-familiar indirect/direct ratio (Flanders, 1970). Yet it has only been in recent years (Cronbach, Gleser, Nanda & Rajaratnam, 1972; McGaw, Wardrop & Bunda, 1972; Medley & Mitzel, 1963) that the area has attracted the attention of measurement specialists. Consequently, greater interest has developed in the measurement properties of observational measures (e.g. Shavelson & Dempsey-Atwood, 1976).

Measurement theory, by and large, has been developed in the context of paper-and-pencil tests of mental abilities and psychological constructs, and many of its concepts and assumptions reflect this origin (e.g. true score, item discrimination, item characteristic curve). The measurement of behavior is an area to which many of these concepts and assumptions have little relevance, and consequently, even the most cherished results of test theory need to be examined before they are applied to behavioral measurements. In this study, the unthinking application of the Spearman-Brown formula to such measurements is called into question, and it is shown that, under certain circumstances, the increase in reliability with increased amount of observation is better described by a quite different formula.

---

[1] Rosenshine (1971), reviewing 51 studies which attempted to relate teaching behavior and student achievement, found many more non-significant than significant correlations reported, and one may probably safely assume that many more non-significant correlations were obtained, but not reported (or, at least, not published). A researcher with information about the reliability of the measures used would have been in a much better position to evaluate the usefulness of such results. In particular, he could judge whether the failure to obtain correlations of the size presumably expected was due to a lack of association between the variables, or simply to unreliability of the measures. This would be a sounder basis upon which to estimate whether further enquiry in the area might prove fruitful.

In setting out criteria to guide the users and developers of observation systems, Herbert and Attridge (1975) included identifying criteria (enabling the user to identify and select the instrument most relevant to his purpose and application), validity criteria (whether the instrument represents accurately and consistently the events of interest), and practicality criteria (ease of use of the instrument, and of the dissemination of results). Given adequate information in relation to these criteria, it is contended that the researcher should be in a position to make a rational choice of an observation instrument.

But can adequate information be provided? Most potential users of an observation instrument will want to know about reliability. Yet it is acknowledged (Herbert & Attridge, 1975, p. 14) that reliability is not a property of an instrument itself, but a property of the measures derived from the instrument. It is clear that the reliability of any measure will depend on many factors other than the instrument used, e.g., the skill of the observer(s), the nature and variability of the subjects of observation, and, particularly, the number and length of the observation periods. The purposes of this study are to uncover the relationships between reliability and the amount of observation undertaken, and to show how these relationships can be used to estimate the amount of observation, and the appropriate scheduling, which will yield an acceptable reliability coefficient.

## Procedure.

The data used in the study were obtained by John Herbert and his associates, as part of a larger study conducted at the Ontario Institute for Studies in Education. The instrument used was a modified version (SAL II) of the System for Analyzing Lessons (Herbert, 1967), and 30 teachers were observed, their behaviors being recorded once per minute for periods of 50 minutes. (Actually, many of the observation sessions extended beyond 50 minutes, but only the first 50 minutes were used in this study). For the 30 teachers observed, the number of 50-minute periods obtained totalled 187, giving, on average 6-7 visits per teacher. Reliabilities were estimated in the manner described by Rowley (1976), yielding both the intraclass correlation coefficient (which may be equated with the expected value of the correlation between scores computed from different observation periods) and the reliability coefficient of the mean score obtained by combining all available observation periods. These coefficients may be interpreted as generalizability coefficients (Cronbach et al, 1972) where the universe of generalization is defined by the way in which the data were collected. If the data were used in a correlational study they would set clear limits to the correlations which could be obtained. Further details are provided by Rowley (1975, 1976).

Three questions were asked of the data:

1. How is reliability affected when the number of observation periods is varied?
2. How is reliability affected when the length of the observation period is varied?
3. How is reliability affected when the number and length of the observation periods are varied simultaneously?

## The effect of varying the number of observation periods.

As n, the number of independent observation periods, is increased, the effect on reliability is well-known, and described in detail elsewhere (e.g. Rowley, 1975; Winer, 1971, p. 286). If $\rho_{11}$ is the reliability of a set of scores obtained from n independent observation periods is given by the generalized Spearman-Brown formula:

$$\rho_{nn} = \frac{n\rho_{11}}{1 + (n-1)\rho_{11}} \qquad (1)$$

Not only is this true of the population values; the method of computation ensures that it will also be true of the sample estimates. For we have seen (Rowley, 1976, p. 54) that $\rho_{11}$ and $\rho_{nn}$ are estimated by the equations:

$$r_{11} = \frac{MS_t - MS_w}{MS_t + (n-1)MS_w}, \qquad (2)$$

$$r_{nn} = \frac{MS_t - MS_w}{MS_t}, \qquad (3)$$

and from these formulae, it is easily demonstrated that

$$r_{nn} = \frac{n\ r_{11}}{1 + (n - 1)\ r_{11}} . \qquad (4)$$

Thus the increase in reliability with increased numbers of observation periods is entirely predictable, and graphs plotting reliability against number of observation periods would all be of similar shape, and identical to the familiar curves obtained when test reliability is plotted against test lengths (see, e.g., Gulliksen, 1950, p. 81, for examples).

The effect of varying the length of the observation periods.

For a fixed number of observation periods, it seems at least intuitively reasonable to expect that the reliability would be greater for observation periods of greater length. The precise nature of the relationship between reliability and length of observation period has not previously been explored. This is an empirical, rather than a theoretical investigation, since no theory presently exists to predict the nature of the relationship.

Rowley (1976) presented reliabilities for 20 measures derived from the same set of data. For each of these 20 variables, one could ask what reliability would have been obtained had each observation period been cut short after 10, 20, 30 or 40 minutes, rather than the 50 minutes that were used. Only a subset of the original variable was chosen for this analysis; since the reliability may be expected to decrease as shorter observation periods are used, only variables with relatively high reliabilities for 50-minute periods were used. Also it was thought desirable that the variables chosen should be of some potential research interest (eliminating variables such as: No Interaction, Neutral Affect, and No Subject Activity). Eight variables were selected, and the reliabilities computed for observation periods of 10, 20, 30, 40, and 50 minutes, using, in each case, every one of the 187 available observation periods to obtain the estimates. The reliabilities thus obtained are displayed in Table 1.

TABLE 1.

RELIABILITIES OF SELECTED VARIABLES, COMPUTED FROM OBSERVATION PERIODS OF VARYING LENGTHS.

| Variable number | Title | Length of observation period (minutes) 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|---|
| | Social Orientation (S) | | | | | |
| 1 | Pupil | .571 | .678 | .748 | .767 | .775 |
| 2 | Whole class | .618 | .677 | .674 | .709 | .760 |
| 3 | Group | .314 | .518 | .580 | .663 | .678 |
| | Nature of Interaction (I) | | | | | |
| 4 | Presenting | .253 | .517 | .702 | .733 | .793 |
| 5 | Watching, listening | .179 | .274 | .473 | .494 | .578 |
| | Affect tone (A) | | | | | |
| 6 | Curt, angry | .682 | .684 | .775 | .809 | .833 |
| 7 | Happy, pleasant | .783 | .870 | .875 | .874 | .857 |
| | Subject Matter (M) | | | | | |
| 8 | Conduct | .381 | .539 | .559 | .631 | .685 |

Close examination of Table 1 reveals that the reliabilities do, in general, increase (as expected) with the length of the observation periods, although the rate of increase is by no means uniform, either among different variables, or within the one variable. Generally, the greatest increment in reliability occurs as the length of the observation period is increased from 10 to 20 minutes, and successively small increments occur as the length increases to 30, 40 and 50 minutes in turn (although not all the variables are equally well-behaved in this respect). This suggests that the relationship of reliability to length of observation period may be mathematically similar to its relationship with the number of observation periods, as described in the preceding section. When the data in Table 1 are graphed, with length of observation period on the horizontal axis, and reliability on the vertical axis, the curves obtained are strikingly similar in shape to graphs of the Spearman-Brown relationship. The lower curve in Figure 1 provides a good example.

The effects of varying simultaneously the number and the length of the observation periods.

We recall that reliability measures the extent to which our set of observations is representative of the wider universe of admissible observations (Cronbach, et al, 1972, p. 20). It seems reasonable, then, to expect that increasing the number of independent samples of behavior would be more effective in producing a representative set of observations than taking the same set of occasions, and observing them for longer periods of time. In the latter case, there is the possibility that the extra observations may contribute very little to the representativeness of the data, particularly if the behavior being observed differs little from that which preceded it.

This explanation proves to be justified, and is illustrated in a specific case by Figure 1. For variable 1, the reliability of scores based upon one ten-minute observation period per teacher was estimated to be 0.176. The reliabilities obtained from two, three, four, and five such observation periods are found by the use of the Spearman-Brown formula: the reliabilities obtained from single observation periods of 10, 20, 30, 40 and 50 minutes are already known. Figure 1 provides a clear demonstration of the general result that, for fixed total observation time, greater reliability is achieved by the use of a larger number of shorter, independent observation periods.

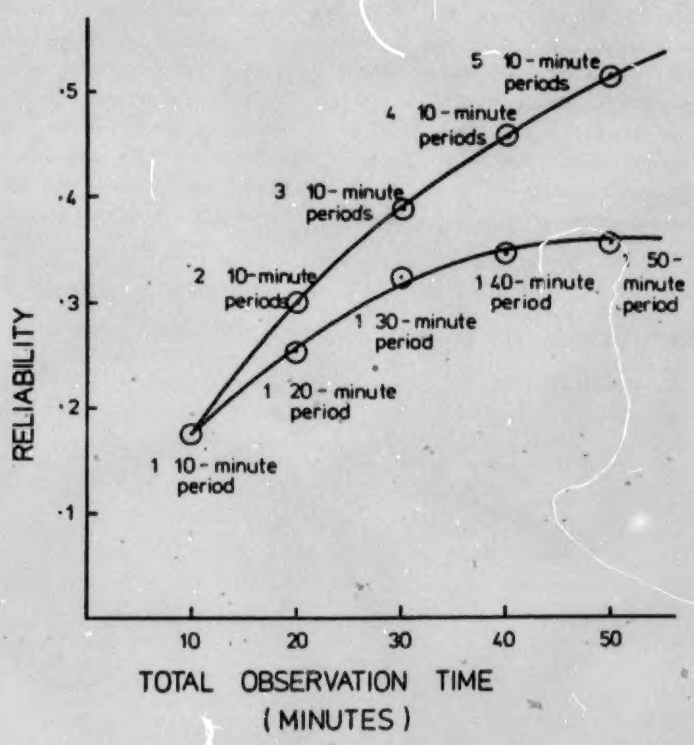


FIGURE 1.

THE EFFECT ON RELIABILITY OF INCREASING FIRST THE NUMBER, THEN THE LENGTH OF THE OBSERVATION PERIODS.

This demonstration is, however, very dependent on the accuracy of one estimate -- the reliability of a ten-minute observation period — and this can be shown (Rowley, 1975, p. 129) to be the least precise of the measurements being used. Table 2, and Figure 2, which is derived from it, illustrate the same general principle in a way which is not so heavily dependent upon the accuracy of that one figure. From Table 2, for example, we may observe that the reliability of 0.176 obtained from one ten-minute visit could be increased to 0.357 by increasing the length of the visits by a factor of five, but to 0.516 by making five times as many visits. In Figure2, the solid curves show the increase in reliability with the lengths of the observation periods for 1, 2, 3, 4 and 5 observation periods, while the broken lines join points which correspond to equal total observation times. The downward slopes of these lines demonstrate the decrease in reliability as the observation periods are made longer and fewer.

TABLE 2.

THE RELIABILITIES OF MEASURE OF VARIABLE 1. AS BOTH THE NUMBER AND THE LENGTH OF THE OBSERVATION PERIODS ARE VARIED

| Number of observation periods | Length of each observation period (in minutes) | | | | |
|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 |
| 1 | .176 | .252 | .323 | .346 | .357 |
| 2 | .299 | .402 | .488 | .514 | .526 |
| 3 | .391 | .502 | .589 | .614 | .624 |
| 4 | .461 | .574 | .656 | .679 | .689 |
| 5 | .516 | .627 | .705 | .726 | .735 |
| 6 | .562 | .669 | .741 | .761 | .769 |

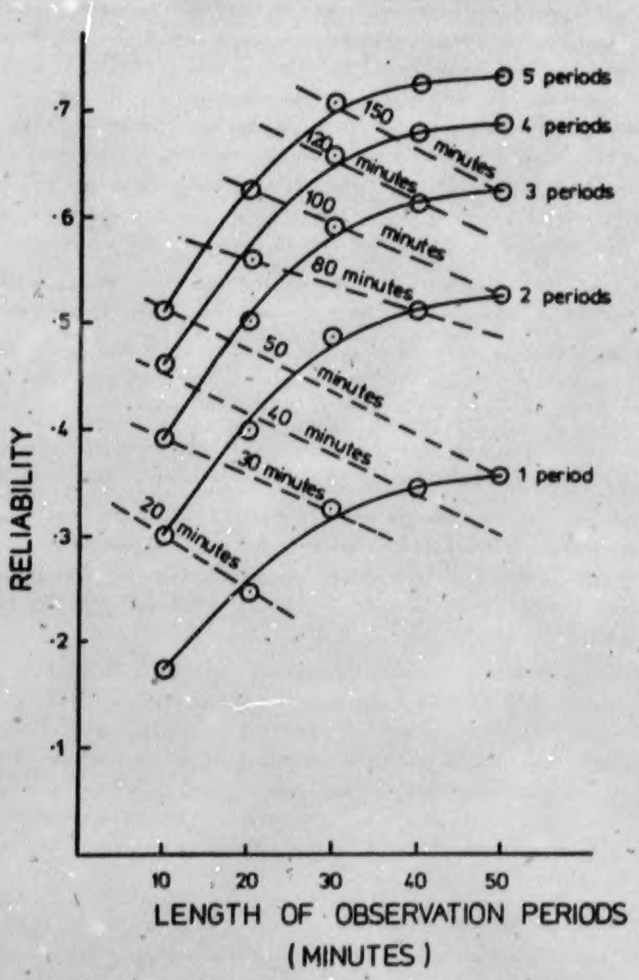


FIGURE 2

THE RELIABILITIES OF MEASURES OF VARIABLE 1, AS BOTH THE NUMBER AND THE LENGTH OF THE OBSERVATION PERIODS ARE VARIED.

These results, which were entirely predictable, are ones to which more heed might be paid in the planning of observational studies. Periods of observation are frequently very long, running, in some cases, to over two hours in this study. Given the high costs of transportation, the desire to achieve the maximum observation time for each mile travelled in understandable. However, it does appear that two hours of consecutive observations would be, at best, only marginally more representative (and hence more reliable) than one hour, or perhaps even half an hour. Had the observation periods been shorter, and more numerous, it seems likely that the observation schedule could have been arranged in such a way as to achieve substantially greater reliability, perhaps even at lower cost.

## Implications for Practitioners of Observational Research.

Classroom researchers have not usually needed to be reminded of the desirability of collecting observational data in large quantities. However, the emphasis on the number of observations has perhaps been misplaced. Were the number of observations the sole criteria of success, the optimal strategy in planning an observational study would be to assign observers to classrooms for the longest practicable time units — perhaps a day at a time, or even longer. Certainly, the effect of this kind of scheduling would be to maximize the number of observations obtained for every mile travelled, and hence for each dollar spent on travel and salaries. However, the empirical results described in this study indicate quite clearly that such scheduling would not maximize reliability; in fact, if reliability be the criterion of effectiveness, such a strategy would be quite wasteful of time and funds. Reliability will be enhanced by a more representative sampling of occasions, and this is best achieved by using a larger number of shorter observation periods. It is not possible to state with any certainty what an optimal length of visit might be; this would depend on many other factors, including the extra travelling incurred by making the visits shorter and more numerous. For example, even if one had reason to believe that maximal reliability would be achieved by observing for only ten minutes at a time and moving on, such a strategy would not "pay off" in any sense if it meant that 99 percent of the observers' time was spent travelling, and only 1 percent was spent observing. Furthermore, one might suspect that the first ten minutes of a class period are somehow unrepresentative of the period as a whole; for example, if they are frequently interrupted by latecomers, or if they are spent checking attendance, or performing other routine chores. If this were even suspected to be the case, it would be more appropriate to sample intact lessons; the reliability might not be greatly enhanced, but the universe of generalization would be of far greater interest.

So many factors affect the reliability of a measure that one cannot expect simple answers to questions such as how best to schedule observation sessions. At best, some general recommendations can be made:

1. Observation periods should be scheduled in such a way as to allow for the maximum possible number of visits to each classroom, consistent with the requirement that each visit should, as far as possible, sample a unit of the class's activities which is complete in itself.
2. Where travel costs are a limiting factor, scheduling should be aimed at increasing the number of classroom visits per mile travelled, rather than just the number of observations for each mile travelled.
3. In any given study, the more uniform the behavior observed during each observation session, the less will be the amount of new information gained by prolonging the visits, and consequently the greater might be the gain from shortening the observation periods and making them more frequent. The more variable the behavior within the observation periods, the less would be the gain from shortening and increasing the number of the observation periods.
4. Within the limitations imposed by points 1 to 3 above, care should be taken to sample occasions as representatively as possible for each classroom. Some strategies, which would otherwise be advantageous, could introduce undesirable biases — e.g., if three of the classes being observed are located in the one school, it may not be desirable to visit them in succession on the same day — at least not as a consistent practice. For if the days chosen were in some sense atypical (the fire alarm went off that morning, examinations started the following day, etc.) the three classes would systematically have been made to appear more similar than they really were. And of course, who could deny that every day, in every school, is in some sense atypical?

## A theoretical rationale for the observed relationships.

That the reliability of a measure should increase with the number of independent observations is well-known, and the reasons are well-documented. If the number of observation periods is held constant, it is intuitively reasonable to expect that the reliability should increase as the length of the observation period is increased, but neither the rationale for nor the precise nature of the relationship have previously been explored. The results presented on this study, suggest that the relationship may be described mathematically by a formula of a type similar to the Spearman-Brown formula.

Suppose we consider a series of $n$ consecutive observation periods, each of the same length, from which measures $X_1$, $X_2$, ... , $X_n$ have been obtained. Let $\rho_{11}$ be the reliability of each of these measures. Since the $n$ observation periods are consecutive, we may combine 2, 3, 4, ... , $n$ of them together to obtain observation periods which are twice, three times, ... , $n$ times as long as the original observation period. Suppose also that, for the longer observation period, we obtain a measure $X$ given by

$$X = \sum_{i=1}^{n} X_i \quad . \tag{5}$$

We recognize that the mean of the $n$ measures may be used in place of the sum; since the reliability would be the same in either case, we will treat only the sum. Also, for reasons of simplicity, and with no loss of generality, we will assume that each measure $X_i$ has been standardized to have unit variance. This assumption is not necessary to the argument which follows, but merely enables it to be presented more simply.

A further, more crucial, assumption is necessary. We recognize, first, the possibility that measurements obtained during the same period of observation are likely to be more similar than measurements obtained from distinct periods of observation. If this were so, the correlation between two of these measures (say $X_i$ and $X_j$) would depend on whether $X_i$ and $X_j$ were obtained from the same or different observation periods. We will propose the simplest possible model which allows for this possibility. We will assume that all correlations between measure $X_i$ and $X_j$ obtained from the same observation period are equal, and that all correlations between measures $X_i$ and $X_j$ obtained from different observation periods are also equal, but possibly different from the first-mentioned correlations.

Formally, we assume that

$$\rho(X_i\ ,\ X_j) = \begin{cases} r\ , & \text{if } X_i \text{ and } X_j \text{ were obtained from different visits.} \\ R\ , & \text{if } X_i \text{ and } X_j \text{ were obtained from the same visit.} \end{cases} \tag{6}$$

We readily identify $r$ with the reliability of the measures $X_i$ $(i = 1, \ldots, n)$ ; $R$ could not be termed a reliability coefficient, since it is a correlation between two experimentally dependent measures. The distinction between $R$ and $r$ corresponds exactly to the distinction between "linked" and independent observations, introduced by Stanley (1967) and described in detail by Cronbach and Furby (1970) in their paper on the measurement of change:

> ... we do not assume that all correlations $\rho_{XY}$ are equal. Sometimes $X$ and $Y$ observations are "linked", as when the two scores are obtained from a single test or battery administered at one sitting, or when observations on different occasions are made by the same observer. The correlation between linked observations will ordinarily be higher than that between independent observations (p. 69).

It is normal to expect that the correlation between linked observations would be higher than that between independent observations (in our case, that $R > r$) because linked observations would contain correlated errors, whereas the errors in independent observations would be independent. As Cronbach and Furby point out (1970, p. 70) it is mathematically possible that the errors in the case of linked observations could be negatively correlated, if the two variables have some complementary relation — "for instance, if rate of reading and comprehension are measured on the same selections simultaneously, one score is likely to rise at the expense of the other, and (the covariance between errors) will be negative" (p. 70). In such a case, the correlation between linked observations would be less than that between independent observations. For the present purpose, we shall assume that no such relationship exists, and that the correlations between errors in the linked case are either positive, at worst, zero, so that

$$R > r\ . \tag{7}$$

The reliability of $X$ would be given by the correlation between two independent measures of $X$, say

$$X = \sum_{i=1}^{n} X_i\ ,$$

and

$$X' = \sum_{j=1}^{n} X'_j\ , \tag{8}$$

where the same assumptions are made concerning the measure $X'$ as were made concerning $X$, in particular that its components $X'_j$ $(j = 1, \ldots, n)$ have the same variances and intercorrelations as those of $X$. Consequently, we find that the reliability of $X$ is given by

$$\rho(X, X') = \rho\Big(\sum_{i=1}^{n} X_i\ ,\ \sum_{j=1}^{n} X'_j\Big)\ . \tag{9}$$

The correlation on the right may be written in terms of variances and covariances:

$$\rho(X, X') = \frac{\sigma(\sum_{i=1}^{n} X_i \; , \; \sum_{j=1}^{n} X'_j)}{\sigma(\sum_{i=1}^{n} X_i) \; \sigma(\sum_{j=1}^{n} X'_j)} \qquad (10)$$

The numerator of this expression may be expanded to give $n^2$ terms, each being a correlation of the "independent" type, and consequently having the value $r$. The denominator reduces to the variance of the measure $X$, which may be expanded to give $n^2$ terms; $n$ of these will be variances, and hence will have the value unity, while the remaining $n(n - 1)$ will be correlations between linked observations, and will have the value $R$. Thus we obtain

$$\rho(X, X') = \frac{n^2 \; r}{n + n \; (n - 1) \; R} \; ,$$

or

$$\rho_{nn} = \frac{n \; r}{1 + (n - 1) \; R} \qquad (11)$$

Equation (11) gives the reliability obtained when an observation period which gives a measure of reliability $r$ is lengthened $n$ times. We note the similarity to the Spearman-Brown formula, and recognize that, in the event that $R = r$ (periods within the same visit are no more similar than periods from different visits), the two formulas would be identical.

The adequacy of equation (11) as a description of the observed relationship may be checked as follows. Suppose that, as we have suggested,

$$\rho_{nn} = \frac{n \; r}{1 + (n - 1) \; R} \; .$$

Then we may take the reciprocals of each side to obtain

$$\frac{1}{\rho_{nn}} = \frac{1}{nr} + (n - 1) \; \frac{R}{nr} \; ,$$

and this yields, with a little re-arrangement

$$\frac{1}{\rho_{nn}} = \left(\frac{1 - R}{r}\right) \frac{1}{n} + \frac{R}{r} \; . \qquad (12)$$

Thus, if the relationship were indeed as described, a graph of $1/\rho_{nn}$ against $1/n$ would yield a straight line.

Rather than examine the data in this way, however, we may proceed to consider the combined effect of increasing both the length and the number of the observation periods. Suppose that, in addition to increasing the length of the observation periods by a factor of $n$, we increase the actual number of observation periods to $N$. We may obtain the reliability of scores based upon the $N$ observation periods by applying the Spearman-Brown formula to the reliability of the scores based upon a single observation period:

$$\rho_{n,N} = \frac{N \; \rho_{n,1}}{1 + (N - 1) \; \rho_{n,1}} \; , \qquad (13)$$

where, in the double subscript, the first subscript indicates the length of the observation period, and the second the number of such periods. The length of the original observation period, before lengthening, is taken as unity.

Substituting for $\rho_{n,1}$, in equation (13) we obtain:

$$\rho_{n,N} = \frac{N \; \dfrac{n \; r}{1 + (n - 1) \; R}}{1 + (N - 1) \; \dfrac{n \; r}{1 + (n - 1) \; R}} \; ,$$

which reduces to

$$\rho_{n,N} = \frac{n \; N \; r}{1 + (n - 1) \; R + n \; (N - 1) \; r} \qquad (14)$$

We note that formula (14) reduces to formula (11) when $N = 1$, and to formula (4) when $n = 1$, as would be expected. We also note that formula (14) may be rearranged to read

$$\frac{1}{\rho_{n,N}} = \frac{(1 - R)}{N r} \frac{1}{n} + 1 + \frac{1}{N} \left(\frac{R}{r} - 1\right) . \tag{15}$$

For any fixed value of $N$, a plot of $1/\rho_{n,N}$ against $1/n$ should yield a straight line, with a gradient given by

$$(1 - R) / Nr , \tag{16}$$

and intercepting the vertical axis at the value

$$1 + \frac{1}{N} \left(\frac{R}{r} - 1\right) . \tag{17}$$

For different values of n, we would expect to find a family of straight lines, with both the gradient and the intercept decreasing as $N$ increases. For variable 1, the appropriate data are provided in table 3, and graphed in Figure 3. The graphs display a strong appearance of linearity, as evidenced by the computed correlation of .997 between $1/r_{nn}$ and $1/n$ (the correlation is the same for each value of $N$ since the Spearman-Brown formula may be shown to be simply a linear transformation applied to the reciprocal of the reliability).

TABLE 3.

DATA USED TO PLOT THE RECIPROCAL OF THE RELIABILITY AGAINST THE RECIPROCAL OF THE LENGTH FACTOR

| $N$ | $n$ | $r_{n,N}$ | $1/n$ | $1/r_{n,N}$ |
|---|---|---|---|---|
| 1 | 1 | .1760 | 1.000 | 5.682 |
|  | 2 | .2517 | .500 | 3.973 |
|  | 3 | .3231 | .333 | 3.095 |
|  | 4 | .3463 | .250 | 2.888 |
|  | 5 | .3566 | .200 | 2.804 |
| 2 | 1 | .2993 | 1.000 | 3.341 |
|  | 2 | .4023 | .500 | 2.486 |
|  | 3 | .4884 | .333 | 2.048 |
|  | 4 | .5144 | .250 | 1.944 |
|  | 5 | .5257 | .200 | 1.902 |
| 3 | 1 | .3905 | 1.000 | 2.561 |
|  | 2 | .5023 | .500 | 2.486 |
|  | 3 | .5888 | .333 | 1.698 |
|  | 4 | .6138 | .250 | 1.629 |
|  | 5 | .6244 | .200 | 1.602 |
| 4 | 1 | .4707 | 1.000 | 2.169 |
|  | 2 | .5736 | .500 | 1.743 |
|  | 3 | .6563 | .333 | 1.524 |
|  | 4 | .6794 | .250 | 1.472 |
|  | 5 | .6891 | .200 | 1.451 |

$r_{n,N}$ is the reliability computed for variable 1, based on $N$ observation periods, each of length $n \times 10$ minutes.

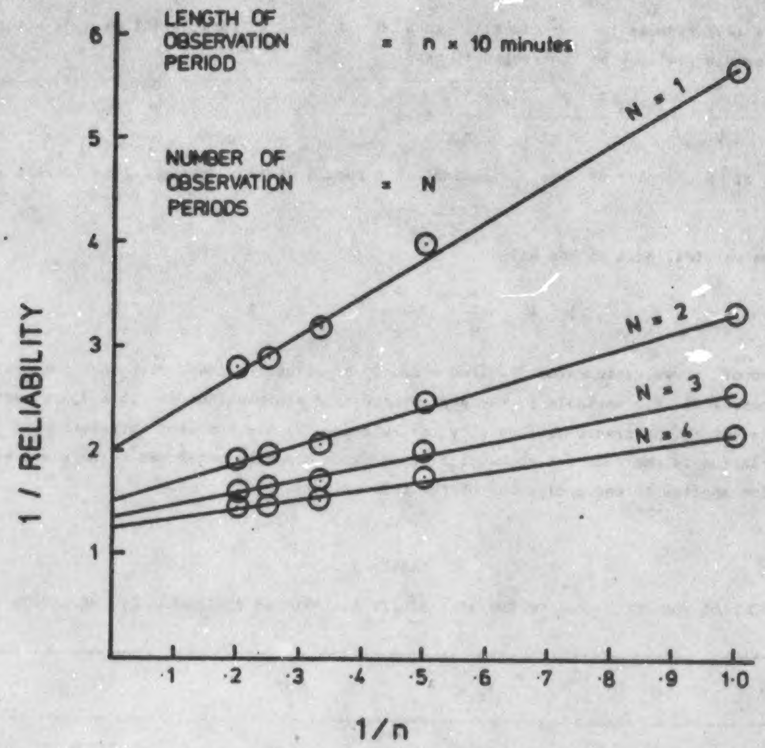


FIGURE 3

GRAPH OF THE RECIPROCAL OF THE RELIABILITY AGAINST THE RECIPROCAL OF n AS THE OBSERVATION PERIODS ARE LENGTHENED n TIMES.

Based on this evidence, then, it seems reasonable to conclude that

1. the rationale provided is a plausible explanation of the relationship which appears to exist between reliability and the number and length of the observation periods, and
2. the formula derived from this rationale reproduces quite accurately at least some of the results obtained empirically.

In support of the second conclusion, the values of $r$ and $R$ were estimated from the data in Table 3 to be .1753 and .3503 respectively.[2] Using these values of $r$ and $R$, we find the predicted relationship to be

$$\frac{1}{\rho_{n,N}} = \left(\frac{3.7065}{N}\right)\frac{1}{n} + \left(1 + \frac{0.9982}{N}\right) . \qquad (18)$$

For $N = 1$, the resulting curve is plotted in Figure 4, along with the experimentally obtained points, so that the accuracy with which the data are reproduced may be judged.

It should be noted, though, that this does not "prove" the correctness of the theory, or of the assumptions behind it, in any general sense. Variable 1 is amongst the best "behaved" variables in this data set; the other variables display what might be perceived as similar relationships, with varying degrees of "scatter". The pursuance of similar analyses with each variable in turn would be a possible next step, but would not contribute very much to our understanding, since the result is quite predictable: greater scatter about the theoretical relationship, as the points scatter more about the smooth curves. It is not claimed, then, that the result is "proven", on the basis of a very good fit from one variable on one data set. Essentially, the theory will predict the experimental results accurately, so long as the data meet the assumptions described by equation (6), and whether this will be true of datasets collected in different circumstances has yet to be determined.

---

[2] A least-squares regression was used to find the best-fitting straight line, and the regression coefficients equated to the expressions in (16) and (17) respectively. The resulting equations were then solved for $r$ and $R$.

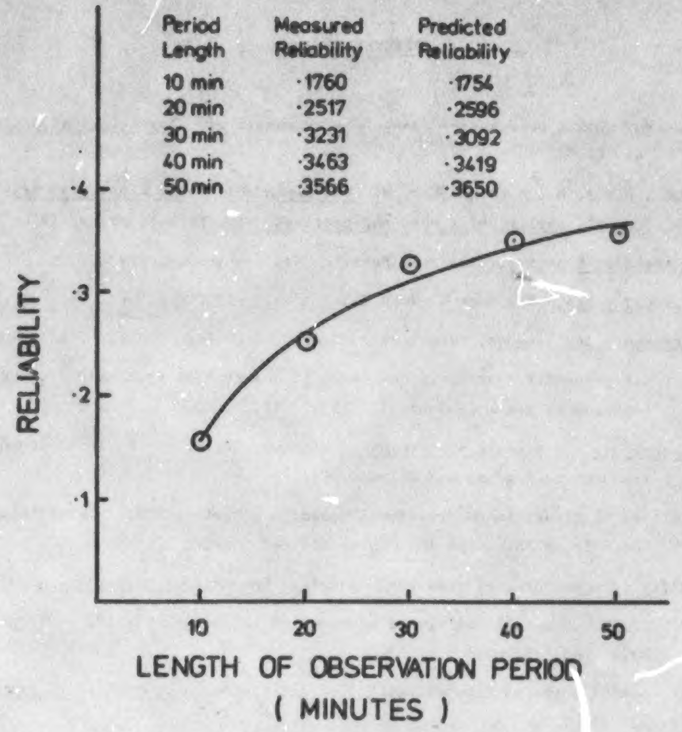


FIGURE 4.

THE RELATIONSHIP BETWEEN RELIABILITY AND LENGTH OF OBSERVATION PERIOD — THEORETICALLY PREDICTED CURVE, AND EXPERIMENTALLY OBTAINED POINTS.

Applications to test reliability.

It may also be worth noting that equation (11), of which the Spearman-Brown formula is a special case, may have wider applications than just those described in this article. For instance, consider a test consisting of N different subtests (e.g., N different topic areas). It may well be that items within the same subtest are more similar than items in different subtests, in which case the correlations between items could be described (approximately, at least) by an equation like equation (6). Since, in practical terms, the only reasonable way of lengthening the test would be by lengthening each of the subtests (rather than producing more subtests), equation (11) may be a more effective predictor of the reliability resulting from an n - fold increase in length than would be the Spearman-Brown formula. This would be the case if the inter-item correlations are indeed higher within subtests than they are between subtests; whether this is generally so, the conditions under which it is likely to be so, and how much difference this makes to the computed reliabilities, are matters which could be the subjects of valuable empirical studies.

## REFERENCES

CRONBACH, L.J. & FURBY, L. How should we measure "change" — or should we? *Psychological Bulletin*, 1970, *74*, 68-80.

CRONBACH, L.J., GLESER, G.C., NANDA, H., & RAJARATNAM, N. *The dependability of behavioral measurements: Generalisability of scores and profiles*, New York : Wiley, 1972.

FLANDERS, N.A. *Analyzing teaching behavior*, Reading, Mass. : Addison-Wesley, 1970.

GULLIKSEN, H. *Theory of mental tests*, New York : John Wiley & Co., 1950.

HERBERT, J.D. *A system for analyzing lessons*, New York : Teachers' College Press, 1967.

HERBERT, J.D. & ATTRIDGE, C. A guide for developers and users of observation systems and manuals. *American Educational Research Journal*, 1975, *12*, 1-20.

ROSENSHINE, B. *Teaching behaviors and student achievement*, Windsor, Berks. : National Foundation for Educational Research in England and Wales, 1971.

ROWLEY, G.L. The reliabilities of classroom observational measures : estimation, interpretation, applications. Unpublished Ph.D. thesis, University of Toronto, 1975.

ROWLEY, G.L. The reliability of observational measures. *American Educational Research Journal*, 1976, *13*, 51-59.

SHAVELSON, R. & DEMPSEY-ATWOOD, N. Generalizability of measures of teaching behavior. *Review of Educational Research*, 1976, *46*, 553-611.

STANLEY, J.C. General and special formulas for reliability of differences. *Journal of Educational Measurement*, 1967, *4*, 249-252.

WINER, B.J. *Statistical principles in experimental design*, (2nd ed.) New York : McGraw-Hill, 1971.

# A SYSTEMS APPROACH TO TITLE I EVALUATION: IMPLICATIONS FOR SMALL CULTURAL GROUPS[1]

Richard B. Baldauf, Jr.
James Cook University of North Queensland

Evaluation, as is fitting for a new area of systematic enquiry and related practices, is still in the process of examining and trialing numerous evaluation models. Worthen and Sanders (1973), Popham (1975), and House (1977) are among those who have described some of the major models and discussed their underlying assumptions. The purpose of this paper is to examine in some detail one of these models the *systems analysis* approach (McLaughlin, 1975). This will be done by summarizing the underlying assumptions of the model, by looking at a specific proposal for its implementation, and by discussing from practical experience problems likely to arise in its use with small cultural groups.

The *systems analysis* model was selected for discussion because "it has served as the major evaluation perspective in the United States Department of Health, Education, and Welfare since about 1965 (House, 1977, 1)." Within the general framework, a subset of particular evaluation strategies was examined since they may soon become the required evaluation and reporting methods for all ESEA (Elementary and Secondary Education Act) Title I projects. These projects, which focus on compensatory instruction in the basic skill areas of reading and math, involve large scale funding of projects in most of the 50 American states and in her Pacific possessions. The potential widespread use of the proposed model makes it one with which evaluators should be familiar.

## A Systems Analysis Approach

A systems analysis approach to evaluation is based on relating quantative output measures, usually test scores, to programme differences (House, 1977). This has often been accomplished through the use of correlational techniques, but more recently has involved the use of experimental design (i.e. Marco, Murphy & Quirk, 1976). The key ideas in the systems analysis model as put forward by its leading proponent Alice M. Rivlin (1971) and as summarized by House (1977, 8) are:

- Key decisions will be made by higher governmental levels.
- The end of evaluation is efficiency in the production of social services.
- The only true knowledge is a production function specifying stable relationships between educational inputs and outputs.
- The only way to such knowledge is through experimental methods and statistical techniques.
- It is possible to agree on goals and on a few output measures.
- There is a direct parallel between production in social services and in manufacturing. The same techniques of analysis will apply.

The aim of the evaluation process in the systems analysis approach is to provide generalizations that will hold in various circumstances. Large samples are required, both to detail the range of circumstances, and to rid the final production function of idiosyncracies. The final product will enable the major consumers of evaluation, the managers and administrators of governmental programmes, to produce social services more effectively (House, 1977).

## Background to the Title I Evaluation System

Title I of ESEA currently provides over US$2 billion annually to American educational agencies, serving approximately 9 million students in areas with concentrations of children from low income families, to provide remedial services in the basic skills. Although the reporting procedures for such grants provide for appropriate objective measurement to be submitted on at least an annual basis, many of the reports submitted at the federal level show a "lack of comparability--- and often of validity --- of the data in them (Anderson, 1977, 2)" The need for better evaluation procedures was further documented by a literature search which

---

[1] A paper presented at the annual meeting of the Australian Association for Research in Education, Canberra, November 1977.

> encompassed some 2,000 projects, all of which had received some form of "official" recognition for success. Of the 2,000, only six could be found which, under close scrutiny, were able to meet the selection criteria of effectiveness, cost, availability and replicability established for this search. Most discouraging, however, was the fact that not one of the evaluations provided acceptable evidence regarding project success or failure. In all cases, problems in conducting and reporting the evaluations rendered the results inconclusive (Horst, Tallmadge & Wood, 1974, 1-2).

In an effort to insure a greater degree of accountability, the United States Congress passed section 151 of the Title I Act in August1974. Under this section the United States Office of Education (USOE) was required to implement a

> *complete* evaluation program: conduct evaluations, upgrade evaluation activities at other administrative levels so that reported data are comparable, use those data to --- among other things --- identify especially effective instructional practices, and disseminate information about those practices. (Anderson, 1977, 3)

Under this legislative mandate, a series of contracts were awarded to the RMC Corporation. The first provided for a review Title I programme evaluations, recommended common reporting practices and checked on the feasibility of the resulting suggestions with a small sample of administrators and evaluators (Gamel, Tallmadge, Wood and Binkley, 1975). The second involved all states and territories and some local school districts in a discussion of a prototype system and its implications for those settings. (Bessey, Rosen, Chian, and Tallmadge, 1976).

From these discussions have emerged three structured evaluation models, with the provision for the use of other models if they can provide comparable data. The system is expected to allow for "the aggregation of unbiased, *project-valid* estimates of the effects of Title I services, ... expressed in one common metric to make such aggregation possible (Anderson, 1977, 11)." Besides providing for aggregation, the system is expected to be useful to identify especially effective educational programmes, to facilitate the monitoring and guidance of programmes, and to help upgrade evaluation activities. "Hence, our hopes for the system are that decision makers at all levels will find benefits from its use (Anderson, 1977, 13)". Local education authorities (LEAs) are expected to begin implementation of these models in the 1979-1980 school year.

## The Evaluation Models

The evaluation models first appeared in October 1974 as an RMC Corporation report. *Measuring Achievement Gains in Educational Projects* (Horst, *et al*, 1974). The Government Printing Office version is reported to have sold over 8000 copies during its first year (Anderson, 1977). Three models based on the above document have been selected and further refined for use with Title I projects.

> ... the models are: Model A, the Norm-referenced Model; Model B, the Control Group Model; and Model C, the Special Regression Model. The Control Group and Special Regression model have variations depending on whether standardized achievement tests or criterion - or objectives - referenced tests are used. The Control Group Model has additional variations in that it may be used with posttest-only comparisons, with or without matching, and with analysis of covariance procedures (Tallmadge & Wood, 1976, 11).

Model A, the Norm-referenced Model is expected to be the one most widely adopted. Pre-post test gains will be compared to test norms. However, the model specifies (a) that standard scores rather than grade-equivalents be used, (b) thattesting be done on dates corresponding to normative points rather than whenever testing is convenient and (c) that pretests be administered *after* participants have been selected (Tallmadge, *et al*, 1976).

Model B, the Control Group Model, requires pre-post test data for treatment and random or "random in effect" control groups. *Post hoc* matching procedures are not allowed, but provisions for analysis of covariance adjustments are made. Model B1 uses norm-referenced tests while Model B2 uses criterion-referenced measures. If model B2 is selected the results are converted to a common metric to make aggregation possible (Tallmadge, *et al*, 1976).

Model C, the Special Regression Model, is the least desirable model because of its potential for misinterpretation. The model incorporates two different evaluation designs, the regression-discontinuality design (Campbell & Stanley, 1963) and the regression-projection design (Horst, *et al*, 1974). Norm referenced (Model C1) or criterion-referenced (Model C2) tests may be used to gather the necessary pre-posttest data (Tallmadge, *et al*, 1976).

It should be clear from the brief description of the models given above "that the system is outcome, rather than process, oriented (Tallmadge, *et al*, 1976, 12)." However the system's originators believe...

> the data called for by the proposed system will do more than provide evidence regarding overall effectiveness of the Title I program. The system will permit analysis of project-level relationships among cost, achievement gains, hours of intervention, grade levels, instructor-pupil ratios, and initial degree of educational need. It will, then, enable investigation of most of the major and minor concerns expressed by educational policy makers interviewed during Phase I of the study (Tallmadge, *et al*, 1976, 12).

## Using the Models with Small Cultural Groups

The work discussed thus far represents an impressive job of synthesizing empirical evaluation methods and detailing models for a wide variety of evaluation needs. Undoubtedly, evaluation has developed to the point where such attempts at synthesis should be encouraged (Gephart, 1977). The mandatory nature of these models will soon be a fact of life for those who want to continue to receive Title I funding. Thus, whether or not one accepts the underlying premises of the systems approach to evaluation as adequate to the understanding of programme functioning, widespread use and dissemination of materials based on this approach is almost certain to occur.

Therefore, the discussion which follows concentrates on problems which are likely to occur when attempts are made to implement and aggregate data using the models. The discussion is limited to use of the models with small cultural groups, specifically in America's Pacific territories, and to problems associated with evaluating English-second-language (ESL) programmes.

Selection of a Model. In the Pacific island setting, where schools tend to be small, communities even within a culture relatively unique, communication is difficult, and where the majority of students only begin to study English when they come to school, the evaluator has few options when it comes to model selection.

Model A, the Norm-referenced Model, is not appropriate for ESL children because standardized tests are normed on native speakers with different linguistic problems and a different rate of language acquisition. Even the Title I norms which are available for some tests are not appropriate as research indicates that ESL children in America's Pacific possessions are considerably more disadvantaged in English language acquisition than, for example, Native Americans (Baldauf, 1977).

Model B, the Control Group Model, is nearly impossible to implement with random or random in effect groups except under limited circumstances in a few of the large junior high and high schools. In general schools are small and not comparable one with the other, and therefore the model is inapplicable.

Model C, the Special Regression Model, is appropriate if criterion referenced ESL tests are used. In American Samoa and the Mariana Islands some appropriate tests are available for this purpose (i.e. Baldauf & Dunn-Rankin, 1973). However, this is not the case (to my knowledge) in the rest of the Trust Territory.

This means that in order to meet the mandatory Title I requirements, LEAs in America's possessions will be forced to use the most complex and least easily interpretable model often with tests that are invalid for the group being examined and the curriculum being evaluated.

Tallmadge and Horst, who have worked with the model development project since its inception, argue that instrumentation "is currently the weakest link in the educational evaluation chain. Until it is strengthened, the information available for policy making decisions will continue to be of marginal utility (1977, 14)." If this is true generally for Title I projects, how valid can evaluations of ESL projects be where there is even less correspondence between instruments and programme content.

At least one attempt is being made to do something about the lack of appropriate tests. In the Mariana Islands, locally designed ESL tests are being psychometrically evaluated, revised and normed on the local population to provide appropriate ESL norm referenced criteria (Baldauf & Annesley, 1977). However, this approach takes several years to implement, under some circumstances may require considerable expense, and demands both expert ESL and evaluation staff familiar with local conditions to produce adequate results.

These requirements provide considerable deterrents to its more general use. Further, there is no indication whether or not locally developed normed-referenced tests will be acceptable in place of Model A in Title I evaluations.

Implementation Problems. Madden (1976) has pointed out some of the potential implementation problems particularly emphasizing the need for training and support programmes by the USOE. The difficulty of implementing the Special Regression Model has already been alluded to. While these aspects are important, Pacific territories have special problems which further complicate proper implementation. These include vast distances, few trained personnel, high staff turnover and higher charges for evaluation materials and services. Their small budgets, based on small numbers of pupils served, make it more difficult to overcome these problems.

There is a danger that the added evaluation demands posed by the implementation of these models will drain away already scarce funds from programme development and implementation. If the models become mandatory, the USOE should seriously consider increasing Title I allocations to Pacific island programmes to ensure the current level of services is at least maintained.

Evaluation Needs. My own experience and that of other ESL programme managers with whom I have talked in the Pacific territories, brings me to the opposite conclusion to those of Anderson (1977) and Tallmadge, *et al* (1976) quoted earlier in this paper with regard to the usefulness of the results for programme managers. The results should be useful for governmental administrators who need to know "if" a programme is working, "for whom", and at what cost. However, the evaluation results will provide little indication of "how" and "why" a programme has succeeded. For the programme manager, the teacher, the researcher, and the evaluator who are concerned about children and advancing our understanding of the learning process, the latter questions are of great importance.

Myer (1975) goes even further than this to suggest that standardized testing, which is about to become mandatory for all Title I programmes, has tended to divert attention from the theoretical problems in reading and lessen the effectiveness of compensatory education.

Thus, while one may sympathesize with the USOE's desire to be able to collect and compile data on programme effectiveness, one cannot help but wish that the data collected would include more formative evaluation which would be of value at the local level for programme improvement. If this is not done, evaluations will continue to be written by evaluators to meet the requirements of other evaluators for the use of distant administrators and statisticians who will be able to claim the successful implementation of social policy. At the local level, evaluations will be neatly filed away, another report completed to obtain funds.

## Summary

This paper has reviewed the systems approach to evaluation and the psychometric models currently proposed to evaluate ESEA Title I projects. The paper suggests that the models may be inappropriate, difficult to correctly implement and detract from programme quality, at least in America's Pacific possessions. The paper recommends allowance be made for a wider variety of evaluation strategies in the final Title I guidelines.

## REFERENCE NOTES

Anderson, J.K. *USOE's Title I evaluation technical assistance efforts: Background and the evaluation system.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Baldauf, R.B., Jr. *The validity of the Michigan Test of English Language Proficience as a general measure of high school achievement in American Samoa.* Submitted for publication.

Baldauf, R.B., Jr. & Annesley, F.R. *Item Analysis Report: Micronesian Test of English Achievement* (MTEA). James Cook University, Townsville, 1977.

Bessey, B.L., Rosen, L.D., Chiang, A. & Tallmadge, G.K., *Further documentation of state ESEA Title I reporting models and their technical assistance requirements.* Mountain View, CA: RMC Research Corporation, 1976.

Gamel, N.N., Tallmadge, G.K., Wood, C.T. & Brinkley, J.L. *State ESEA Title I reports: Review and analysis of past reports, and development of a model reporting system and format.* Mountain View CA: RMC Research Corporation, October, 1975.

Gephart, W.J. *Evaluation reconsidered: Do we need a synthesis? Definitely!* Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

House, E.R. *Assumptions underlying evaluation models.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Madden, J.V. *Summary of comments to the policy advisory committee on the proposed Title I evaluation models*, Mimeo, March 1976.

Tallmadge, G.K. & Horst, D.P. *Using the data from state and local ESEA Title I Reports.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Tallmadge, G.K. & Wood, C.T. *Users Guide.* Mountain View, CA: RMC Research Corporation, 1976 (draft).

## REFERENCES

Baldauf, R.B. Jr. & Dunn-Rankin, P. *A progress report on a plan for curriculum assessment in American Samoa.* Pago Pago: Department of Education, 1973 (ERIC ED 074 200).

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental design for research on teaching. In N.L. Gaye (Ed.), *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963.

Horst, D.P., Tallmadge, G.K. & Wood, C.T. *Measuring achievement gains in educational projects.* Los Altos, CA: RMC Research Corporation, October 1974 (UR-243). Also *A practical guide to measuring project impact on student achievement.* Washington, D.C.: U.S. Government Printing Office, 1975. (017-080-01460).

McLaughlin, M.W. *Evaluation and reform.* Cambridge, Mass.: Ballinger Pub. Co.: 1975.

Marco, G.L., Murphy, R.T., & Quirk, T.J. A classification of methods of using student data to assess school effectiveness, *Journal of Educational Measurement*, 1976, *13*, 243-252.

Myers, M. Uncle Sam's a reading puppeteer: Guess who the puppet is? *Learning*, 1975, Nov., 20-23, 26.

Popham, W.J. *Educational evaluation.* Englewood Cliffs, N.J.: Prentice Hall, 1975.

Rivlin, A.M. *Systematic thinking for social action.* Washington, D.C.: Bookings Institute, 1971.

Worthen, B.R. & Sanders, J.R. *Educational evaluation: Theory and practice.* Worthington, Ohio: Charles A. Jones, 1973.

Richard B. Baldauf, Jr. is Lecturer in Education at James Cook University of North Queensland. He has a Ph.D. in Educational Psychology from the University of Hawaii, and has taught English as a second language for three years in Sabah, Malaysia. He was Supervisor of Educational Testing for three years in American Samoa, and has been involved in Title I and Title III project evaluations in American Samoa and the Mariana Islands. His research interests include measurement and evaluation problems in the Pacific region, with an emphasis on English second language learning and cultural change.

This research was supported in part by a University Research Grant from James Cook University of North Queensland and by funds provided by the U.S.O.E. to the Mariana Islands under ESEA Title III. The opinions stated in this paper are those of the author and no endorsement of the work by the U.S.O.E. is implied.

# DESIGNING POLICY ORIENTED EDUCATIONAL RESEARCH

**Neil Baumgart and Brian Low**

## Introduction

This paper is written at an early stage in our thinking on the topic and we would be grateful for reactions and criticisms. We feel strongly motivated towards the topic for a variety of reasons. These reasons stem partly from what we see as our professional responsibility and partly from our own philosophical stance.

Two kinds of frustrations voiced in recent years have made us aware of the lack of communication between policy-makers and researchers and of the need for more research in education to be policy oriented. On the one hand, those responsible for educational policy formation and decision-making at the system level have expressed disappointment that researchers have failed to deliver the kind of research findings relevant to crucial decisions. Martin Shipman, who left the academic world of universities to head the research division of the Inner London Education Authority, expressed this view most strongly in his opening address to the 1976 Annual Conference of the British Education Research Association. On the other hand, some educational researchers feel frustrated that, for whatever reason, their research seems to have so little impact on important educational policies and decisions. Of course, this is probably not a significant problem for those researchers committed to basic research and to the pursuit of knowledge *per se*. But for those of us who believe our research should have some short term utility and should also contribute to directions for future development, these frustrations are very real.

The two kinds of frustrations both stem from an apparent inability of educational research to contribute to major policy issues. In a paper originally presented as an invitational address to the 1974 AARE Conference in Melbourne, Start (1975) suggested a number of reasons for this. He reminded us that policy decisions are often taken not on solely educational grounds and that social, political or economic criteria can be dominant. He further suggested that educational research is frequently *ineffectual* for policy makers because

(i) the time scale is inappropriate. Researchers frequently respond to immediate problems and neglect likely future problems;

(ii) solutions implied by research findings are not always feasible in practical terms;

(iii) research results are often equivocal. Researchers present findings appropriately qualified in terms of a study's limitations and hence clear guidelines for policy decisions are difficult to infer;

(iv) research results are seldom disseminated in a format likely to be heeded by "key" policy-makers;

(v) researchers experience conflict between actively recommending implementation strategies based on research findings and keeping an open mind as a prerequisite to further research.

Start (1975, p. 336) advocated, among other things, "a much closer dialogue between the researcher and the policy-maker". We also saw this as essential and set ourselves the task of devising a venue and a format to facilitate such communication for a particular set of policy decisions.

We confine our attention to *major* policy decisions taken at the *system* level. By major decisions, we mean those having implications for large numbers of people (say pupils or teachers or parents) or those involving expenditure of large amounts of public money (say, more than $10,000). By system level, we have in mind State or Territory Departments of Education although we believe that our approach could be applied equally well in other systems.

Our objectives for the present paper are as follows:

(i) to define key terms related to the topic;

(ii) to explore the current and potential relationships between policy decisions and research information relevant to those decisions;

(iii) to develop a simple taxonomy of policy decisions;

(iv) to develop a similar taxonomy of information relevant to policy decisions;

(v) to identify research methodologies relevant to (iv);

(vi) to consider the interface of (iii) and (iv); and

(vii) to outline objectives for a proposed "study" on policy oriented research within the domain outlined above.

## Terms

A focus on educational decisions has occurred both in the research literature and in the evaluation literature in recent years. Thus the often cited report of the Committee on Educational Research of the National Academy of Education (Cronbach and Suppes, 1969) drew the distinction between decision-oriented and conclusion-oriented research. Similarly, several evaluation models with an emphasis on decision-management (for example, Alkin, 1969; Stufflebeam et al., 1971) have received widespread attention.

An educational decision requires a choice from a set of discrete and competing alternatives. Thus within a State system of education, decisions are made on whether or not to reduce class size, whether or not to moderate (or abandon) a matriculation examination, the number of teaching scholarships to be awarded, and so on. As noted by Birch (1976, p. 1), "the term 'policy' is an elusive one". As used in this paper, the term educational policy is both broader in scope and less explicit in its specification than the universal set of educational decisions. Thus

> Policies within an organizational setting may be viewed as generalizations that involve the goals of the organization, as well as the ways in which goals can be implemented. Policies affect the structural properties of organizations, the directions they take, and the way in which resources are allocated". (Spuck et al., 1975, p. 212)

To put the two terms together as we have done above and to speak of a policy decision implies at least an identification of discernible alternatives within the policy requiring a deliberate choice on the part of the decision-maker. It means that, within the general policy framework, particular issues are made explicit to permit action or a statement of intended action.

Policy research or policy oriented research seeks to collect, analyze and report information that will facilitate the making of policy decisions and/or the formulation of policy statements. Note that the attempt here is to provide research findings to assist, or in some cases to influence, the decision-maker. It should not be confused with research on how policy is made, or what the political scientists term policy analysis.

It is appropriate to note here the now widespread use of the term "policy sciences" as distinct from the physical and social sciences and the rapid expansion in the literature associated with this. The boundaries of this literature are not easy to define and it tends to cross a number of different areas of study. A good deal of the literature overlaps with that variously described as futurism, futory, futuristics, and so on.[1] Lonsdale (1975) distinguished policy research from future research although he drew attention to their close inter-linking characteristics.

## The Role of Policy Research

The relationships between research findings and policy decisions are most complex. Even where educational policies do rely heavily on research information - and this may be a rare event - it would be naive to assume a simple one-to-one relationship between the information and the policy. Start (1976, p. 257) discussed four categories for research before suggesting that "perhaps policy research is to social sciences what applied research is to the sciences, technologies and engineering".

Using a familiar systems approach, Craven (1975, p. 128) distinguished decision making within an organization on three levels - "operations, management, and planning". He defined these and their relationships to policy as follows:

> Operational activities are those focusing primarily on routine, day-to-day detailed transactions within an organization... Management activities are those that provide for the general control of routine operations. Planning activities are directed at the short- and long-range development of program and policy direction of the institution. With risk of overgeneralizing, operational activities effect policy, management activities monitor the implementation of policy, and planning activities establish policy.

Craven warned that the boundaries between these types of decisions were often blurred but contended that a significant implication of the classification was that different kinds of information were required by different types of decisions. Thus information required for operations was likely to be "routine transactions" typically produced as "routine reports, usually in a fixed format". At the next level, information necessary for management decisions was less detailed and was usually derived from "management analysis and special reports". Craven saw this information used "for monitoring, co-ordinating or controlling operational functions, for responding to ad hoc requests ... in planning, and for providing reports to outside agencies". At the planning level where policy statements are formulated, the information needed to be in a format that summarized broad trends rather than gave precise details.

---

[1] Lonsdale (1975, p. 251) gives 20 different labels. At least one volume on the topic has been published by an Australian University (Emery, 1974).

In as much as it provides a factual data base, educational research may contribute to the management and even to the operations functions of an education system or organization. What Start (1976) referred to as the "optimising" role for educational research would seem to be of this type. Given that a policy decision has been taken, research can aid management in making its implementation as efficient and as effective as possible. Decision-management evaluation models (for example, as proposed by Alkin, 1969 or by Stufflebeam et al., 1971) would also seem to fit here. Formative evaluation models at least overcome the problem of time lag referred to above.

However, the major potential contribution of policy oriented educational research, and one as yet unrealized, is towards planning functions and their attendant formation of policies. It is at this level that writers express different views about the role of educational research. One possibility is to view educational research as having a service function. The onus is on the policy-maker to spell out policy needs; the onus is on the researcher to meet these needs and to convey findings to the policy-maker in a mode which is understood and in a time frame relevant to policy decisions. Start (1976, p. 262) described the relationship as follows:

> ... policy research usually stems from a policy need and that has to be communicated to the researcher by the policy-maker. Such a route should not be exclusive naturally, but it would be the main one. Hence there must be a dissemination of "policy" to those capable of undertaking the research.

While this may be a natural relationship when policy-makers and researchers work within the system, the service function of policy research could occur in a broader framework. Thus Start (1976) has anticipated future expansion in the client-contractor roles for policy-makers and researchers respectively with more emphasis on commissioned research. He considered other possibilities in which public discourse occurred or policy-maker and researcher exchanged ideas but saw these limited by the "high political vulnerability" of proposed policy if it were "peddled in the open market place".

Other writers have questioned whether this is a viable role for policy research. Cohen and Garet (1975, p. 19-20) claimed that such a role for policy research needs to make two assumptions - firstly, "that policies consist of discrete decisions"; and secondly, "that applied research is more authoritative than ordinary common sense ideas about social policy". They question the validity of these assumptions, particularly in education. They see a policy, not as a set of decisions with clear alternatives, but "as a grand story: a large and loose set of ideas about how society works, why it goes wrong, and how it can be set right" (p. 21).

On the basis of their analysis, Cohen and Garet concluded that policy oriented research is likely to influence "broad assumptions and beliefs underlying policies" rather than particular decisions. Hence they claimed that policy research needs to be seen as a discourse between researcher and policy-maker in which both parties can initiate, set rules, express views, request clarification, question, and so on.

In similar vein, Etzioni (1971, p. 8) claimed that policy research ought not to "take on specific assignments" but should maintain a critical stance. In this respect, he claimed policy research differed from applied research "in much the same way that strategy differs from tactics; it is more encompassing, longer-run in its perspectives, and more concerned with the goals of the unit for which it is undertaken". Iannaccone (1975, p. 232) adopted a similar view and claimed that policy research needed a critical component rather than being simply solution oriented. Thus the policy researcher could provide alternative approaches to problems and had the responsibility for challenging the assumptions of policy-makers.

In our view, the relationship between policy-maker and researcher in education is, and is likely to remain, both complex and variable. The extent to which policy research is solution oriented on the one hand or critical on the other will vary according to the nature of the policy, the decision setting, the social and political climate, and so on. Whatever the relationship, the need for adequate communication between policy-maker and researcher is evident. Yet it is probably just this ingredient which is most elusive. In our view, it is this absence of dialogue which is at the heart of the two kinds of frustrations we discussed initially.

Concerned about this state of affairs, we decided to embark on a study in which one of the major purposes would be to initiate some form of dialogue between policy-makers at the system level and researchers. We saw as a necessary first step the need to develop a taxonomy for major educational policy decisions. We considered it would be more productive at this stage to focus on identifiable decisions by policy-makers in which alternatives were considered and some rejected although we acknowledge that these might only be indicators of broader policies as conceived by Cohen and Garet (1975).

Since the major purpose of the taxonomy was to facilitate communication, simple categories were preferred to complex ones even if it proved difficult to classify some policy decisions. Furthermore, we decided that communication should be facilitated if initially we developed separate taxonomies for

(i) the policy decisions; and
(ii) the information relevant to those decisions.

Apart from facilitating communication, to be able to describe the interface between these two categories for different classes of policy decisions was in itself a legitimate research objective.

These two taxonomies are now described briefly. They are by no means our first attempt but nor do we see them as our last. Hence we welcome suggestions from Conference members.

Taxonomy of Policy Decisions

Figure 1 represents a two dimensional taxonomy for education policy decisions.

SUBSTANTIVE CONTENT

TIME FOCUS

| | O | C | F | M | PP | PR |
|---|---|---|---|---|---|---|
| HISTORICAL | | | | | | |
| CURRENT | | | | | | |
| SHORT-TERM | | | | | | |
| LONG-TERM | | | | | | |

CODE
O: ORGANIZATIONAL
C: CURRICULAR
F: FISCAL
M: MANPOWER
PP: PHYSICAL PLANT
PR: PUBLIC RELATIONS

FIG. 1 TWO DIMENSIONAL TAXONOMY FOR POLICY DECISIONS

The first dimension is labelled substantive content of the decision and has six categories - organizational, curricular, fiscal, manpower, physical plant, and public relations. We see the first category as essentially structural and the remaining five as functional. Our preliminary trials suggest that the categories are mutually exclusive but their exhaustiveness has not been adequately tested to date.

The second dimension is a time one. The time refers to when the action or intended action occurs, not to when the decision is taken. Initially, with an emphasis on planning, we thought only in terms of two categories - short and long term. Operationally, we define short term as occurring within the next two years; and long term as occurring beyond that time.[2] We then extended the time scale to include current time in an attempt to accommodate decisions based on formative evaluation strategies. What Start (1976) describes as "band-aid" decisions also fit here. Finally, we also decided to include an historical category to meet the cases where research findings are used to justify actions taken previously. Summative evaluation can be of this kind although (hopefully) it will carry implications for future action even if this future action is to maintain the status quo.

We considered these two dimensions - a substantive one and a time one - were central to a policy decision. For a particular policy decision, we would hope to classify the decision on these two dimensions so that it can be located unambiguously within one cell of the matrix in Figure 1. Of course, a policy decision also has many other characteristics - it is made by a particular person or group, it occurs within a particular setting, and so on. We preferred to treat these factors as subsidiary ones and examine them briefly later in the paper.

---

[2] Some writers and planners would include a category for middle range projections and forecasts as well.

## Taxonomy of Information Relevant to Policy Decisions

Again we propose a two dimensional classification of information (see Figure 2) used in policy decisions.

NATURE OF INFORMATION

| DATA SOURCE | CW | DI | P/T | P-R | A/C |
|---|---|---|---|---|---|
| ECONOMIC Ec | | | | | |
| EDUCATIONAL Ed | | | | | |
| POLITICAL P | | | | | |
| SOCIAL S | | | | | |

CODE

CW: CONVENTIONAL WISDOM
DI: DESCRIPTIVE INFORMATION
P/T: PROJECTIONS AND TRENDS
P-R: PREDICTIONS BASED ON RELATIONSHIPS
A/C: ALTERNATIVES AND THEIR CONSEQUENCES

FIG. 2 TWO DIMENSIONAL TAXONOMY FOR INFORMATION RELEVANT TO POLICY DECISIONS

The first dimension - nature of the information - has five categories. The term "conventional wisdom", borrowed from Goodlad (1966), is used to describe the common sense accumulated from past experience. The second category provides descriptive information on an existing state of affairs or program and a means of feedback. In the third category, projections and trends, we see essentially one variable charted over time and then extrapolated to yield future estimates. The fourth category, predictions based on relationships, is intended to include information derived from research studies to indicate relationships among sets of variables. Such studies may be experimental but it is more likely that they will be large scale, multivariate and correlational in nature. To be useful to policy decisions, key variables need to be subject to manipulation within the system. Finally, the fifth category is labelled alternatives and their consequences. Some of the alternatives will have been posited by the policy-maker but we see this category as open-ended to permit inclusion of information on possible alternatives not originally considered by policy-makers.[3]

For this dimension, the five categories are not necessarily mutually exclusive. In some instances, they may be seen as hierarchical and cumulative.

The second dimension included here has been labelled data source with four categories - economic, educational, political, and social. Start (1975) discussed "criteria" for policy decisions under these labels and indicated that on many occasions they may conflict. He further suggested that educational outcomes often had to take fourth place behind political and economic feasibility and social implications. While recognising the existence of such priorities, we considered it would be very difficult to establish directly the relative importance of different criteria underlying a policy decision. For this reason, we chose to regard these as data sources available to a policy-maker and hence we include this dimension to describe information.

For a particular policy decision, these four categories are not mutually exclusive. That is, a policy-maker may seek information from any or all of the four sources. Suppose we code this information on Figure 2 as follows:

+ Information considered; favours the alternative selected
- Information considered; contrary to the alternative selected.

---

[3] In deciding which alternatives to consider, or, for that matter, in deciding which variables to include in a study of relationships, value questions are inherent. Instead of trying to avoid these value questions and to be objective bystanders, researchers may well be advised to make value issues explicit. See Cooley and Lohnes (1976) for a discussion of this point of view.

Blank Information not considered from this source.

Thus, a particular policy decision coded on the two information dimensions may yield a complex pattern. Figure 3 is an example.

NATURE OF INFORMATION

DATA SOURCE

| | CW | DI | P/T | P-R | A/C |
|---|---|---|---|---|---|
| Ec | | | - | | |
| Ed | | | + | + | |
| P | | | - | | |
| S | + | | | | |

CODE

+ CONSIDERED; FAVOURS ALTERNATIVE SELECTED

- CONSIDERED; CONTRARY TO ALTERNATIVE SELECTED

FIG. 3 EXAMPLE OF PATTERN OF INFORMATION USED IN A PARTICULAR DECISION

The pattern conveys information on the kinds and sources of data used and, just as importantly, on the kinds of data not used. Moreover, if the pattern contains some positives and some negatives, inferences may be drawn about the relative importance of different information sources. Thus, for the example coded in Figure 3, the policy decision was based on favourable educational predictions and was in line with commonsense social views in spite of adverse economic and political predictions. Of course, the example is fictitious.

As in the case of policy decisions, information can have other characteristics - for example, analyses to render it more intelligible, a mode for communication, a format and a time for presentation, and so on. Again, these are not included in the taxonomy but need to be considered for different classes of decisions.

## Information and Research Methodologies

The taxonomy of information discussed above used five categories for nature of the information. Although we consider a combination of methodologies is likely to yield the best information in any instance, the different categories do suggest some methodologies as more appropriate than others. Table 1 is an attempt to summarize some possible methodological approaches under the different information categories.

Many of the methodologies listed in Table 1 have been applied in disciplines other than education. Apart from the contribution educational researchers should be able to make in terms of providing relevant, timely information to policy-makers, there is also a need for some researchers to contribute to the methodology of policy research. This may involve the application in new contexts of techniques such as those listed. But it may also involve the modification of these strategies to meet new circumstances or the development of new strategies.

Table I. A Classification of Methodologies[4] according to Information Type

| Conventional Wisdom | Description and Feedback | Projections and trends | Predictions based on relationships | Alternatives and their consequences |
|---|---|---|---|---|
| | Most evaluation models (see Worthen and Sanders, 1973, p.210)<br>Flow models<br>Network analysis techniques (e.g. PERT[5], PATTERN[6]) | Exploratory forecasting<br>Projection models<br>Trend analysis<br>Time series analysis<br>Growth analysis<br>Cohort analysis<br>Markov chains | Modeling or simulation<br>Monte Carlo techniques<br>Path analysis<br>Evaluation research models using canonical analysis as proposed by Cooley and Lohnes (1976)<br>(Delphi techniques) | Normative forecasting<br>Force analysis<br>Cross-impact matrix<br>Bayesian models<br>Contextual mapping<br>Morphological analysis<br>Ariole<br>The Scenario |

## The Interface of Policy Decisions and Information

So far we have considered a taxonomy for policy decisions in which a particular decision can be located in one cell of the matrix in Figure 1. Then we considered a taxonomy of information in which a particular policy decision called for a pattern of coding across the matrix of Figure 2. In bringing the two together, we wish to move beyond a particular decision and to search for general patterns. Are some kinds of decisions more likely to be made using certain kinds of information from selected data sources? Do some of the cells from the policy decision matrix (Figure 1) form clusters with similar informational properties?[7]

These are themselves research questions we hope to pursue by considering a set of policy decisions taken at the system level. Of course, if policy-makers decide on appropriate alternatives and seek research findings in relation to particular problems via contractual agreement with researchers, then the kind of analysis envisaged here is of little import. But if, on the other hand, policy-makers and researchers can establish the kind of dialogue envisaged by Cohen and Garet (1975), then identification of general patterns can facilitate such dialogue. It can guide researchers on the kinds of information most likely to be used by policy-makers in respect of classes of decisions. For policy-makers, it can point to additional kinds and sources of information pertinent to crucial decisions.

As anticipated above, if general patterns can be established between policy decisions and information requirements, then additional descriptive data would also be useful in respect of each pattern. Such description might include

* the identity of key decision-makers
* the decision setting
* what analyses of data assist interpretation of findings
* a suitable format for reporting information
* a means of dissemination of information to decision-makers
* a timetable for particular decisions.

## A Proposed Research Study

We expressed above a particular interest in policy decisions at system level. We hope to identify key decision-makers in State and Territory Departments of Education and to interview them re previous major policy decisions and likely future ones. With the policy decision as the unit of analysis, we then propose to

(i) identify classes of policy decisions with similar informational requirements;
(ii) relate classes of policy decisions to the kinds and sources of information most likely to be used by policy-makers;

---

4. See Hencley and Yates (1974); Lonsdale, (1975)
5. Program Evaluation and Review Technique.
6. Planning Assistance through Technical Evaluation of Relevance Numbers.
7. At the time of writing, we have only made some preliminary explorations with simulated data on methods of quantitative analysis relevant to this problem.

(iii) describe other characteristics of each class of policy decision.

We then propose to send this information, appropriately summarized, to educational researchers (say, to AARE members) and to seek their responses to perceived information needs. From these responses, we hope to

(i) identify information needs not being met by educational researchers;
(ii) identify potential information sources not being tapped by policy-makers.

In then providing a report to the policy-makers who participated initially, we would hope to initiate a form of dialogue. We have given some thought to how such dialogue could be maintained but this would depend on the success of our initial attempt to establish communication. Perhaps an ERDC based Clearinghouse with a modest commitment to on-going staff would suffice. Or perhaps a journal seeking contributions from researchers and policy-makers on policy research at this level would be useful.

## Summary

The major concern of this paper has been to seek ways to improve communication between policy-makers and researchers. We have attempted to classify policy-decisions on the one hand and research information relevant to them on the other as a prelude to exploring the interface between them. We see such exploration as an area of needed research and propose one study to investigate this at system level. In this way, we hope to

* indicate to researchers "productive" areas of research yielding information likely to be used in policy decisions;
* indicate to policy-makers potential areas of research knowledge relevant to policy decisions;
* initiate a form of dialogue which could have a continuing role in the improvement of both policy making and policy research.

## References

Alkin, M.C. (1969). Evaluation theory development. Evaluation Comment, 2 (1), 2-7.

Birch, I.K.F. (1976). Forming Australia's Education Policy. Melbourne: Macmillan.

Cohen, D.K. and Garet, M.S. (1975). Reforming Educational Policy with Applied Social Research. Harvard Educational Review, 45 (1), 17-43.

Craven, E.C. (1975). Information Decision Systems in Higher Education. Journal of Higher Education, 46, (2), 125-139.

Cronbach, L.J. and Suppes, P. (eds.)(1969). Research for Tomorrow's Schools; Disciplined Inquiry for Education. N.Y., Macmillan.

Emery, F. (1974). Futures We're In. Canberra: A.N.U.: Centre for Continuing Education.

Etzioni, A. (1971). Policy research. American Sociologist, 6, 8-12.

Goodlad, J.I. (1966). School, Curriculum and the Individual. Waltham, Mass.: Blaisdell.

Hencley, S.P. and Yates, J.R. (eds.)(1974). Futurism in Education: Methodologies. Berkeley: McCutchan.

Iannaccone, L. (1975). The Field Study in Educational Policy Research. Education and Urban Society, 7, (3), 220-238.

Lonsdale, R.C. (1975). Futures Research, Policy Research and the Policy Sciences. Education and Urban Society, 7, (3), 246-293.

Spuck, D.W., Hubert, L.J., and Lufler, H.S., Jr., (1975). An Introduction to Educational Policy Research. Education and Urban Society, 7, (3), 211-219.

Start, K.B. (1975). Reality for the Researcher. American Educational Research Journal, 12, (3), 323-336.

Start, K.B. (1976). Policy Orientated Educational Research. New Directions in Australian Education. Melbourne: The Australian College of Education.

Stufflebeam, D.I., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.I., Merriman, H.O., and Provus, M.M. (1971). Educational Evaluation and Decision Making. Icasca, Illinois: Peacock.

Worthen, B.R. and Sanders, J.R. (1973). Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones.

## FRIDAY 11 NOVEMBER

## 2.00 - 3.00pm

| PAPERS | Rooms |
|---|---|
| OMITTED<br>J. Maling-Keepes (W.A.I.T.)<br>Educational Evaluation: Key Characteristics | Seminar Room A |
| R. Lewis (La Trobe University)<br>Evaluation of Teachers' Classroom Behaviour: Some Methodological Considerations | Ladies Drawing Room |
| J.C. Clift, B.W. Imrie (Victoria University of Wellington)<br>A Model of Behaviour for Evaluating Courses in Tertiary Institutions | Seminar Room D |
| T. Williams, P. McDonnell (Australian National University<br>Public and Private: School System Effects on Social and Economic Attainments in Australia | Meeting Room |
| R. Straton (University of Sydney)<br>Research of the Evaluation Process: Current Status and Future Directions | Gateway Room |
| OMITTED<br>TAFE Workshop (2.00-4.00pm)<br>Chairman, R. Catts, N.S.W. Department of TAFE)<br>J. Malley (R.M.I.T.)<br>A Case Study of Needs Evaluation<br>R. Schilling (S.A. Dept. F.E.)<br>A Case Study of Needs Evaluation<br>D. Kuhl (S.A. Dept. F.E.)<br>Course Evaluation - Rationale<br>P. Edwards (N.S.W. P.T.C.)<br>Discussion Opener | Seminar Room B |

# EVALUATION OF TEACHERS' CLASSROOM BEHAVIOUR:

## SOME METHODOLOGICAL CONSIDERATIONS

Ramon Lewis

La Trobe University

A question basic to a paper entitled "Evaluation of Teachers' Classroom Behaviour", is 'what criterion does one use for the evaluation?' Many criteria are possible, for example the consistency of a behaviour with a teacher's personal philosophy of education. However, a search of the relevant literature shows that one criterion has received far greater attention than any of the others. This is the teacher's effectiveness at producing student achievement of curricular goals. Most of the goals described have been cognitive, with student attitudes apparently being considered less significant and/or less researchable. Therefore, for the sake of simplicity, the discussion in this paper will focus upon the class of studies which investigate the effect of teacher behaviour on student achievement.

Although the title of the paper mentions teachers' behaviour it is probably adviseable to explain that this term is intended to include those behaviours operationally defined by 'low inference' measures of teaching acts (for example 'asks questions' or 'gives directions') as well as 'high inference' assessment of characteristics of teacher style or classroom climate (for example 'well organised' or 'clear in presentation'). Those measures are usually obtained from external observers or the teacher's students, who report on teaching behaviour. No distinctions will be drawn between the different forms of operational definitions of teacher behaviour in this paper, as the discussion is applicable to them all.

Most of the studies which have focussed on the teacher's classroom behaviour and its subsequent effect on curricular outcomes can be accurately described as decision-oriented (Cronbach and Suppes, 1969). That is, these studies of classroom interaction have been primarily concerned with identifying effective or superior patterns of teaching behaviour, so that trainee and practising teachers could be helped to adopt these patterns under instruction. This approach may be contrasted with conclusion oriented research which is primarily concerned with determining relationships between variables for purposes of describing the real world in theoretical terms. It is within the decision-oriented framework of research that some of the methodological issues confronting the classroom interaction researcher can be usefully considered.

### Definition of the Treatment Variable.

The first methodological issue to be treated will be the manner in which the teacher behaviour is defined. Investigations of the effect of teachers' behaviour were originally undertaken by experimenters who, it may be argued, wished to demonstrate the value of democratic, as opposed to autocratic, leadership. Thus the initial definitions of teacher control were largely a reflection of the philosophical standpoint of the researcher and had little empirical foundation.

The use of an empirically unjustified composite variable has been a major weakness in the research designs employed by many of the early studies investigating teacher behaviour. This criticism is not new. It was voiced more than 20 years ago by McKeachie (1954) in his review of studies of teachers' styles, and by Rosenborough (1935) with respect to cases of more general group leadership situations, and again for both of these types of situations by Anderson (1959). A reason that appears to have been influential in establishing the common use of a composite definition of teacher behaviour was the philosophical bias of early experimenters against authoritarian aspects of teaching.

> "When teacher control was branded as autocracy and contrasted with so-called democratic teaching, all the research seemed to be trying to prove that teacher control cannot possibly work...Almost without exception the dominative leaders are taught to be harsh, to use hostile comments in criticising students and to discourage student initiative".
>
> (Cronbach, 1954, pp. 455-456)

In the light of this comment and the other early criticism of the compounding of the teacher control variable, it may be surprising that as recently at 1974, Dunkin and Biddle found it necessary to continue to criticise studies for the combination of two apparently independent dimensions of warmth and directiveness. However, there are numerous attractions about more compound definitions. A multi-faceted and hence 'fuller' description appears to facilitate the location of a teacher on a dimension. The measurement by observation of teachers in their classes is a costly and difficult task, and the wider type of definition seems to be one way of facilitating it. Secondly, teaching is made up of a great many interactions between teachers and students. To atomise it too finely may obscure the essential characteristics of a relationship between two human beings over the period of time of a particular educational task.

In more recent years, with the advent of appropriate statistical techniques, compositely defined teacher behaviour measures have been based on criteria other than philosophical bias. The accepted procedure in recent investigations has been to describe groups of teaching behaviours which occur together naturally, and then examine their effects on the student's achievement. In general, the statistical procedure utilised for grouping these naturally co-varying behaviours has been Factor Analysis. For example, Anderson (1959) refers to the analyses of Gekoski (1952) and Carter (1952). In addition, Solomon, Bezdek and Rosenberg (1964), after describing the results of their factor analysis, compare their dimensions of teacher behaviours to those found in six other factor analytical studies to which they refer.

The statistical procedure called Factor Analysis however, can only play a 'necessary but not sufficient' role in defining teacher treatment variables. The components of the composite definitions produced in the factor analytical studies have all been defined as logically distinct behaviours and although there is evidence of association between them in classrooms, this does not provide the justification that is required for the use of the sum of their scores to measure an effective independent variable in a decision oriented piece of classroom interaction research.

It has been suggested earlier that the aim of research into the effects of teacher behaviour is to identify effective modes of teaching. Stated more statistically, the aim is to maximise the prediction of some curricular outcomes that can be attributed to specific teacher behaviours.[1] If the decision-oriented emphasis of classroom interaction research is accepted then the first consideration in combining component behaviours should be their tendency to relate to student achievement in a similar manner (for example, both positively) and for each to contribute some independent explanatory power. Assume for a moment that a behaviour called 'emphasis on factual knowledge' affected student achievement negatively and 'use of formal planning' had an opposing positive effect. It would be illogical to investigate the effect of the sum of these two behaviours, despite their tendency to co-vary in the natural situation. To do so involves the unstated assumption that it is impossible to break down the relationship between these two aspects of teacher style and manipulate one while the other remains unchanged. Furthermore, to report on a total score would seem to indicate a belief that the nature of the relationship will remain unchanged even when, through manipulation, the separate teacher behaviours may be exaggerated beyond the values which occurred in the naturalistic settings of the investigation.

On first consideration an assumption of continuing dependence, despite manipulation, appears reasonable. If teachers who tend to display behaviour X also display Y, then changing X could be assumed to similarly affect a change in Y. The disturbing thing about this assumption is that practically all of the experimental studies of the effects of teacher behaviour on student achievement appear to be based on the opposite assumption. They assume that teachers' teaching behaviours are distinct and can be manipulated independently, thereby allowing the effect of component behaviours to be determined. This position appears to be supported in part by the many Competency Based Teacher Education (C.B.T.E.) programs. These programs developed from the micro teaching (Allen and Fortune, 1966) and mini course (Borg et al, 1970) procedures which involve the independent acquisition of specific teaching skills and which appear to have been quite successful.

> "In summary, the microteaching and mini-course programs and other C.B.T.E. efforts have shown that teacher process behaviours can be taught."
>
> (Good, Biddle and Brophy, 1975: 35)

However, the assumption of independence of teaching behaviours also has its problems. Rosenshine and Furst when speaking of Church's study (1971) state

> "One of the clearest findings is that one cannot vary a behaviour such as 'open questions' and expect the rest of the interaction to remain constant."
>
> (Travers, 1973: 124)

---

[1] The appropriate statistical procedure for such determination would be multiple regression analysis, but the models to be used will not be discussed. They can be found elsewhere (Lewis 1976).

The most logical and acceptable solution would be to assume that some relationships between teacher behaviours may be manipulated whereas others are permanent and are not amenable to outside influences. Empirical investigations would be required to determine the permanence of relationships between teacher behaviours. At this relatively primitive stage, however, rather than assume inherent dependence, which may prevent the discovery of potentially significant relationships between teacher behaviour and student achievement, it would probably be wiser to assume them independent until proven otherwise.

The need for caution before the use of a composite variable is accepted is indicated by a few studies which have reported separate findings for single, low inference teaching behaviours, and which have found, for example, that the giving of directions is positively correlated with achievement, while the use of criticism correlates negatively (Rosenshine, 1971). Combination of these two variables, a practice inherent in the use of Flanders' very popular $^{I}/_{D}$ ratio (1965), is clearly likely to reduce the chance of finding significant main effects.

It would be desirable and most convincing if further substantial evidence could be presented to indicate the conflicting effect on criteria of different aspects of teacher behaviour commonly used to form one composite variable. However, such data are not available. Accordingly, in the face of the abundance of studies with non-significant findings, the potential that compound variables have as a contributor to this situation is justifiably stressed until much more convincing evidence is available that the separate components do interact supportively on the outcome variables of interest.

Measurement of the Treatment Variable.

A second design problem confronting those researching the effects of teacher behaviour on student achievement stems from the existence of Aptitude-Treatment Interaction and its implications for the manner in which the teacher treatment' should be measured. Aptitude-treatment interaction is a term used to denote the idea that there may not be a 'best' teacher treatment for all students, but rather, certain types of teacher behaviour may be effective for some students and not others. The effect of these interactions have been the subject of several comprehensive reviews (Bracht, 1970, Berliner and Cohen, 1973, and Tobias, 1976). The implications of the findings of these reviews for classroom interaction research are that it is not sufficient to look at main effects of teachers on their students overall. There should be an exploration of which of his or her students a teacher affects. A similar interaction that is not frequently considered is the possible effect that the teacher's personality may have on what teaching behaviours are effective. It may therefore be necessary to also consider for which teacher a particular behaviour is 'effective'. However, the decision oriented framework of investigations appears to indicate a belief that not all 'effective' teaching behaviours are moderated by the teacher's personality and that there are some 'intrinsically' effective behaviours of benefit to all teachers.

In order to consider the practical consequences of a desire to measure the effect of teacher behaviour on individual students, it is useful to consider the teacher's behaviour (or style) as a 'Press' on his or her students (Murray, 1938). It can then be subdivided into Subjective press and Objective press.

The Objective press represents the objective reality of the situation as defined by the judgements of disinterested, trained external observers. Subjective press results either from the action of each student's own perceptual filter on the 'real teaching' - providing a possibly distorted image - or from differential treatments meted out by the teacher to particular students. In considering the significance of the two presses for educational research, the following arguments need to be noted.

If it is the aim of the researcher to relate his findings regarding 'effective' teacher behaviour to the training of teachers, then the teacher style defined by the researcher should be manipulatable. For example, if the classroom research was aimed at identifying the most effective type of teacher behaviour so that it could be taught to trainee teachers, then it is important to assess teacher behaviours that are manifest.

On the other hand, regardless of the situation that really exists, it is inevitable that the student's perceptual filter will colour the real situation. The resulting personalised interpretation of teacher behaviour may or may not be closely related to reality. Despite its lack of external objective validity, it is nevertheless the student's interpretation that acts as the stimulus which results in whatever outcomes are recorded. To monitor the real teacher therefore is not necessarily going to provide a record of this influential variable. A desire to monitor the student perceptions naturally restricts the classroom interaction researcher to the use of individual student ratings.

The distinction between these two presses should not be ignored. The measurement of Subjective press necessitates the use of student ratings. Objective press appears to require an external (objective) observer. However, if a study involved many teachers over extended classroom lessons, or if dyadic interactions between the teacher and individual students in the classroom are to be recorded, the cost in time and manpower for external observers becomes prohibitive. One solution to this problem of obtaining a measure of the Objective press is to use the class average of student ratings. Gardner (1972) refers to several studies which support his argument that this measure of a teacher's behaviour is concordant with that obtained from an external observer. In this manner 'consensual press', defined by the class mean of student ratings can be utilized as an approximation of the Objective press. Therefore, if the range of the Subjective presses in a class is small, they can logically be ignored and all analyses can be done with the Objective press.

One technique used to determine the extent of Subjective press in previous research is to partition the total variance of students' perceptions of teacher style into the variance within class and the variance between class means. This analysis provides an indication of the relative strength of students' digression in their perception of the same teacher,from the consensus of opinion of classes on their different teachers. An analysis of variance comparing within class variance to between class variance indicates to what extent the variation in perception of teacher behaviour is attributable to class membership. If the extent of Subjective press warrants its consideration then separate analyses need to be performed.

These analyses could either take the form of an Analysis of variance or co-variance nested within classrooms, or a regression analysis with each student being allotted the class mean for the teacher treatment variable. However, the statistical implications will not be described in detail here, but can be found in Lewis (1976).

## Significance of Findings.

The term 'significance' is one frequently used by researchers when describing effects and relationships. In many cases the term is assumed to imply that which dictionary definitions attribute to it; meaningfulness and importance. However, in the context of educational research the word has a more technical and limited use, namely the statistical definition of significance. This is operationalised for particular sets of data by means of tests which indicate if relationships between measured variables exist to a greater extent than might be expected due to chance. These tests provide an estimate of the confidence that can be placed in the generalisability of relationships found in samples.

Most of the studies which have focussed on the teacher's classroom behaviour and its subsequent effect on curricular outcomes have utilized this traditional statistical concept of significance. Researchers have often reported 'significant' findings, concluding that particular teaching behaviours lead to gains in learning outcomes, without considering if the statistical concept of significance is appropriate. To determine the appropriateness of a statistically significant teacher effect requires an examination of its relevance to the previously stated purpose of classroom interaction research.

The problem can be satisfactorily illustrated with two examples. Tuckman demonstrated a very statistically significant relationship ($p < 0.001$), between Teacher Directiveness and Student Satisfaction (1968, p. 56, Table 9). As a result of this study he made the following recommendation.

> Teachers in training should be given more information about the non-directive teaching style and should be afforded more supervised opportunity to practise it. Since this study showed that the non-directive style was more effective (i.e. more satisfying, more preferred, and yielding higher grades), teachers would do well to develop a capability to use this style, particularly if it is not one that comes to a particular teacher naturally. p. 72.

Tuckman's recommendation could be accepted by teacher training institutions as an invitation to alter curricula. However, it is not necessarily sufficient to attribute meaningfulness to certain teacher behaviours on the basis of the mere existence of statistically significant relationships. A precise measure of the extent of the relationship is vital, since the decision makers would need to justify differential 'Cost' in time and money by reference to the likely greater achievement of the curriculum outcomes as a result of implementation of alternatives.

Statistical techniques are available for such determinations. For example, the size of an effect can be found by manipulation of the sum of squares figures in traditional analysis of variance or co-variance tables. This approach provides an estimate of the amount of the criterion variance (or criterion variance adjusted for co-variates) that can be uniquely attributed to each of the independent variables. What then is the significance of Tuckman's data? Cohen (1968) provides the formula 'Sum of Squares of treatment divided by the Total Sum of Squares' for determining the size of treatment effects in analysis of co-variance tables. Application of this formula to Tuckman's data provides a value of 0.029. This means that, although the mean satisfaction scores of students with directive and non-directive teachers differed by a statistically significant amount, the percentage of the satisfaction variance that can be related to teaching style is only 2.9 percent.

A more recent and well known investigation into teacher styles by Bennett (1976) reports

> The effect of teaching style is statistically and educationally significant in all attainment areas tested (p. 152).

Application of Cohen's formula to Bennett's data (Table 2 in Appendix C) indicates that initial achievement in reading 'explains' or 'predicts' 68 percent of the variance of the post-test reading scores. However, teaching style accounts for 0.52 percent of variance of the post-test scores adjusted for pre-test scores. Clearly 0.52 percent of explained variance, despite its reported statistical significance, is not what one may call educationally significant.

Similar calculations based on Bennett's data (Tables 3 and 4, in Appendix C) indicate that teaching style is related to 0.89 percent of the adjusted mathematics scores and 0.63 percent of the adjusted English scores.

It would appear that both Tuckman and Bennett have assumed that their results are not only statistically significant but are also educationally meaningful. A number of other examples could be cited to demonstrate the assumption by research workers that statistical significance implies educational significance. However, low percentages of the criterion variance associated with the variable under study raise considerable doubts about the justification of this assumption.

What then makes a finding concerning a relationship between an independent variable and a criterion measure educationally significant? In the absence of detailed knowledge about the situation of those who may be interested to adopt the finding, some arbitrary level would have to be chosen. Where the situations are better known, consideration of the 'cost-benefit' for each group of potential participants, that is, their likely gain in relation to the cost of the planned change can assist in determining what will be regarded as educationally significant. This will differ for each group of participants. For example, certain 'productive' teacher behaviours may be more cheaply incorporated by practising teachers than by the curriculum planners in teacher training institutions. Each of these interested parties is more likely to be able to respond to information that refers to the size of the 'productive' relationship rather than the probability of its existence.

As indicated earlier, the present level of sophistication of statistical knowledge allows for relatively clear indications of the sizes of effects between teaching behaviours and student achievement. For example, it is possible to determine the percent of variance of the achievement scores that can be related to specific teaching behaviours, or the magnitude of Beta coefficients in regression equations may be calculated to allow determination of the change in student achievement that would be related to changes in teacher behaviour. Therefore accurate evaluations of the value of the findings of previous investigations could feasibly be calculated. What is disturbing and disappointing, however, is the continued production of texts such as 'Teachers make a Difference' (Good, Biddle and Brophy, 1975). These books tend to discuss the statistical difficulties inherent in interpreting findings in some detail, yet nevertheless manage to slide unjustifiably from talk of 'statistical' significance to that of 'practical' significance without any attempts at interpreting the magnitude of relationships.

## The Need for Reassessment of Previous Research.

It is not my contention that teachers do not make a difference, but rather than the difference they make has not been demonstrated as sufficiently convincing enough to warrant the sorts of conclusions consistently being made. Summaries of previous research could, and probably will, result in the identification of teaching behaviours of practical significance. However what is required is an analysis of the literature on the effects of teacher behaviour on the scale recommended by Glass in "Primary, Secondary and meta-Analysis of Research".

> "Those who manage financial support of educational research and evaluation must realize that the integration of research findings is a scholarly undertaking requiring costly literature searches, extensive data analyses and time to reflect and write."
>
> (Glass 1976: 4)

Glass considers, among several comprehensive reviews of research, Dunkin and Biddle's 1974 review "The Study of Teaching". He is critical of the methodology employed and refers to the 'Voting method' of tabulating significant and non-significant findings, without consideration of differing design attributes of the studies (size of sample, level of schooling of subjects, etc).

> "It is not suited to the task of answering the important questions of how large an effect a particular treatment produces...".
>
> (Glass 1976: 6)

In his example of a process called 'Meta analysis', Glass classifies studies of psychotherapy outcome research by important design characteristics. He then reanalyses their data to allow a comparative presentation of the size of treatment effects in terms of the standard deviation of the outcome measure, describing the average member of the treatment group in terms of the percentile rank of members of the untreated population.

## Usefulness of the accepted Research Paradigm.

The last methodological consideration relates to the model upon which research into the evaluation of teachers' classroom behaviour has generally been based. Traditionally, investigators have operationally defined 'objective' data and collected such data in a way that facilitates the application of standard statistical procedures. This paradigm necessitates statements of teachers' behaviours and characteristics prior to the research being conducted. There is generally no attempt to determine what the teacher perceives as his or her actions, nor what teachers believe their role should be. Similarly, because of the 'objective' nature of the design of the investigation there is little, if any, opportunity for obtaining measures of teacher or student variables which the subjects of the investigation may have subjectively felt to be intrinsic to the teaching-learning process, but which may not have been considered in the planning of the study. To provide for the discovery of subjectively significant student and teacher attributes, which may be of assistance in the interpretation of teacher-student interactions, an additional research paradigm must also be considered. This is the 'Illuminative' model proposed by Parlett and Hamilton (1972).

The illuminative paradigm involves the use of what are considered anthropological research techniques; for example, participant observation, taking field notes, recording and in depth interviewing. Delamont and Hamilton (1976) argue very strongly the need for the acceptance and use of this additional research paradigm, citing among their evidence the relatively small contributions to understanding that have accrued from the masses of research that have investigated the affect of Teacher behaviour. Delamont (1976) in a chapter entitled "Beyond Flander's Fields" provides an example of the way in which the two paradigms can be jointly used to enhance the quality of the researcher's findings, by allowing the Illuminative model to add 'flesh to the bones' provided by the traditional model.

Despite Delamont and Hamilton's tendency to blame the lack of consistency and significance of findings of classroom research on the use of one inappropriate paradigm, their hypothesis has yet to be tested. Clearly, the value of the traditional paradigm cannot be determined until it is adequately represented. In this paper some very basic issues pertaining to the valid implementation of the traditional paradigm have been raised. These included reference to the definition and measurement of teachers' behaviour and the manner in which results should be interpreted and summarised. Until these methodological weaknesses in classroom interaction research are confronted and overcome, very little evidence of the value of the traditional paradigm will continue to be available.

## REFERENCES

ALLEN, D. & FORTUNE, J. An analysis of micro-teaching: New procedures in teacher education. In Micro-teaching: A Description, Stanford: Stanford University Press, 1966.

ANDERSON, R.C. Learning in Discussions: A resume of the authoritarian democratic studies. Harvard Educational Review, 1959, 29, 201-215.

BENNETT, N. Teaching Styles and Pupil Progress, Open Books, London, 1976.

BERLINER, D.C. & COHEN, L.S. Trait-treatment interaction and learning. Review of Research in Education, 1973 1, 58-94.

BORG, W., KELLY, M., LANGER, P., & GALL, M. The Mini-Course: A Micro-Teaching Approach to Teacher Education, Beverly Hills: MacMillan Educational Services, 1970.

BRACHT, G.H. Experimental factors related to aptitude-treatment interactions. Review of Educational Research, 1970 40, 627-645.

COHEN, J. Multiple regression as a general data analytic system. Psychological Bulletin, 1968, 70, 426-443.

CRONBACH, L.J. Educational Psychology, New York: Harcourt, Brace & World, 1954.

CRONBACH, L.J. & SUPPES, P. Research for Tomorrow's Schools: Disciplined Inquiry for Education, New York: MacMillan, 1969.

DUNKIN, M.J. & BIDDLE, B.J. The Study of Teaching, 1974, Holt, Rinehart & Winston, Inc.

FLANDERS, N.A. Teacher Influence, Pupil Attitudes and Achievement, Co-operative Research Monograph No. 12, U.S. Government Printing Office, Washington, D.C., 1965.

GARDNER, P.L. Attitudes to Physics, Unpublished Ph.D. Thesis, Monash University, 1972.

GLASS, G.V. Primary, secondary and meta-analysis of research. Educational Researcher, 1976, 5(10), 3-8.

DELAMONT, S. Beyond Flanders fields: The relationship of subject-matter and individuality to classroom style. In Explorations in Classroom Observation, Stubbs, M., & Delamont, S. (eds.), John Wiley & Sons, 1976.

GOOD, T.L., BIDDLE, B.J. & BROPHY, J.E. Teachers Make a Difference, Holt, Rinehart & Winston, 1975.

DELAMONT, S. & HAMILTON, D. Classroom research: A critique and a new approach. In Explorations in Classroom Observation, Stubbs, M., & Delamont, S. (eds.) John Wiley & Sons, 1976.

LEWIS, R. *The Effect of Selected Teacher and Student Behaviour on the Cognitive and Affective Objectives of PSSC Physics*, Unpublished Ph.D. Thesis, Monash University, 1976.

MURRAY, H.A. *Explorations in Personality*, New York, Oxford University Press, 1938.

PARLETT, M.R. & HAMILTON, D. *Evaluation as Illumination: A New Approach to the Study of Innovatory Programs*, Occasional Paper 9, Centre for Research in the Educational Sciences, University of Edinburgh, 1972.

ROSENSHINE, B. *Teaching Behaviours and Student Achievement*, London, National Foundation for Educational Research, 1971.

SOLOMON, D. BEZDEK, W.E. & ROSENBERG, L. Dimensions of teacher behaviour. *Journal of Experimental Psychology*, 1964, *33*, 23-40.

TOBIAS, S. Achievement treatment interactions. *Review of Educational Research*, 1976, 46(1), 61-74.

TRAVERS, R.M.W. (ed.) *Second Handbook of Research on Teaching*, Rand McNally, College Publishing Company, 1973.

TUCKMAN, B.W. A study of the effectiveness of directive versus non-directive vocational teachers as a function of student characteristics and course format. (U.S. Office of Education Research Project - No. 6-2300) Rutgers University, 1968.

# A MODEL OF BEHAVIOUR FOR EVALUATING COURSES IN TERTIARY INSTITUTIONS

J. C. Clift and B. W. Imrie
Victoria University of Wellington

## INTRODUCTION

This paper is not concerned with the specific techniques of evaluation - the means of exploiting the accessible sources of information about the teaching-learning processes. These techniques and their deployment are variously described by Clark, *et al.* (1977), Dressel (1976), Goldschmid (1977), and Miller (1977).

The authors are concerned that the evaluation of courses in tertiary institutions should not have a confusion of motives, and should have a model of behaviour for the benefit of all concerned. So-called teaching-learning processes will be subsumed more implicitly as interactions between teachers, subjects and students. These interactions are conditioned, of course, by the nature of the course, the department, the institution, and the society, not to mention (explicitly) the people involved - all of them random variables.

Goldschmid (1977), in his recent review, *The Evaluation and Improvement of Teaching in Higher Education*, which comprehensively includes references from the U.S.A. and Canada, Europe and the U.K., Australia and New Zealand, cites 122 references, two of them prior to

1970, with the evaluation explosion indicated in Figure 1. This is not to imply that prior to 1970 the field of evaluation was barren; in fact, there is considerable literature, but its emphasis was towards measuring academic attainment of the students or comparing educational programmes in terms of the quality of student output.

---

FIGURE 1

---

It was Cronbach (1963) who, in his paper, "Course Improvement through Evaluation", attacked this over-emphasis on the use of evaluation for making decisions about students and teachers, and about the quality of one programme as compared to another. To these two roles he added a third: "the collection and use of information aimed at improving the quality of an educational programme".

While Universities are currently debating ways and means of measuring the teaching effectiveness of individual lecturers for purposes of promotion and appointment, Higher Education Units, now established in the majority of Australian and New Zealand Universities, are primarily concerned with Cronbach's third category - selling the idea to staff that they should be using evaluation in its role as an aid to improving the quality of their courses. It is true that such an evaluation will seek information about the lecturers' teaching style and perceived effectiveness, but this is but one component of this category of evaluation. Just as important is the seeking of information about student behaviour and how this has interacted with lecturers' style, the content of the course, and the context in which the course is occurring.

A recent survey, "Current use of Evaluation" (1976), in which staff in a Science Faculty were asked to indicate their use of course evaluation, produced data which provided a description of the typical evaluation carried out in large classes.

Information is gathered by questionnaires designed in the form of a rating scale, enabling computer processing. The questionnaire is handed to students in the last week of the course in a one-shot approach, with the return rate of completed questionnaires often less than 50%. Many of these evaluations are conducted with no real goal in mind except maybe feelings which are expressed by such phrases as:

> "It is a valuable means of political appeasement."
>
> "It permits students to feel more like participants than victims."

Occasionally lecturers will admit that an evaluation enables them to modify and improve a course.

Many lecturers, having used a standard computer statistical package for the analysis of their 50-100-item questionnaires, on receiving their print-out suffer "statistical indigestion", the effect of which has been described by Katz (1950) as "mental dazzle".

Such factors:

(a) the lateness of the exercise;

(b) the one-shot approach;

(c) the use of questionnaires with little attempt at validation;

(d) the restricted sample;

(e) the lack of real objectives;

(f) the attitude "it is the thing to do";

(g) the failure to share information with others;

have produced many studies of doubtful quality. It is little wonder that many lecturers lack a certain confidence in the reliability and validity of course evaluation, and that Goldschmid (1977) talks of the lack of impact of efforts to improve teaching in higher education. In fact, the one-shot end-of-course questionnaire can be counter-productive for many lecturers seeking professional development.

For those working in Higher Education Centres the problem is a very real one, as many believe that evaluation is a major if not vital technique in their efforts to help staff improve the quality of their courses and the experiences of their students.

How can the impact of evaluation studies be improved?

Firstly, by helping lecturers to design evaluation programmes that:

(a) have clearly defined 'research' questions;

(b) place minimal workloads on staff and students.

Secondly, but no less important, establishing what Goldschmid (1977) has called a consultation-evaluation approach which provides for:

(a) involvement of both staff and students;

(b) a 'sharing' of information between all the parties involved;

(c) a 'context' in which change, if necessary, is acceptable;

(d) regular feedback to all parties involved;

(e) a multi-procedural approach.

## A MODEL OF PEDAGOGICAL EVALUATION

The model used by the University Teaching and Research Centre at Victoria University of Wellington in working with staff in conducting evaluation studies, is based on the principle that an evaluation study is a learning situation in that it helps the lecturer and student to learn about their course. The model emphasises the active and interactive role of course members in the evaluation and, in particular, their responsibility for the evaluation tasks. Centre staff work with people, not for them.

A basic premise is that, in the learning situation, regular feedback is important. In this regard, the conclusions reached by Thorndike and reported by McKeachie (1976) are considered to be of extreme importance, namely that the feedback must provide meaningful information; the receiver must be at least interested in improving, and there must be better alternatives available to him. If evaluation aimed at improving the quality of a course can be described as a learning situation, then we can use a learning model to seek guidance as to the factors that should be considered in the design, implementation, reporting, and decision-making stages of an evaluation.

Another important premise in the Victoria University of Wellington model is that, for educational research (evaluation) to produce results from which recommendations stem and which, when actioned, produce positive change, there needs to exist a "receptive climate". McGehee (1957) believes that the failure of training research to solve training problems was due to what he called the "general training climate". Courses only occur in a context and, if recommendations are to be of value, they must be administratively practicable. Schutz and Silberman (1966) consider public relations as an essential component, and Dressel and Lorimer (1961) consider that a study that does not take into account the complex inter-relationship existing in an institution is doomed to failure. These "stone walls" can only be breached by sound interpersonal communication.

Let us then consider a model which sees the evaluation as a learning situation taking place within a "receptive climate" created and maintained by communication and active involvement of all parties concerned.

While we cannot observe learning directly, we can describe the events that lead to learning. These events can be classified into three stages: an apprehending stage, an acquisition stage, and a remembering stage (Di Vesta, *et al.*, 1920). For the purpose of the evaluation model, the remembering stage is seen as the stage at which the findings are applied to the solution of problems. In the proposed model is is therefore called the "application stage".

## Apprehending Stage

In an evaluation study, the first stage is of particular relevance. A lecturer enters the apprehending stage from a variety of sources (Figure 2). The Centre's initial tasks are to respond and to focus the lecturer's attention so that an effective evaluation study can proceed. During the apprehending stage, the objectives of the study need to be debated and agreed upon. Then can follow discussions on such matters as the techniques and procedures to be followed, the tasks to be undertaken, the extent of the involvement of staff and students, and the likely effect of this upon their existing workload. Full discussions of these matters are necessary if lecturer anxiety is to be reduced and an attitude of co-operation developed. It is at this stage - the gaining and controlling of interest - that a sound rapport needs to be established amongst all those involved.

---

FIGURE 2

---

Often the initial request is phrased in a very general and vague manner.

> "Our reasons for seeking as objective an evaluation as methods permit are based essentially on the need for planning curriculum development in the Department."

At the initial meeting, the Centre's representative attempts to interpret the meaning of an "evaluation" study in an endeavour to establish a "frame of reference" in which discussions can then take place. It is often appropriate for the lecturing staff to be asked to read a report of a previous evaluation and to discuss the implications with their colleagues involved in the reported study. In developing this procedure, the Centre, in addition to reporting summaries of data and comment from which recommendations have been derived, is endeavouring to provide reports in which emphasis is given to procedures, staff commitment and attitudes.

With involvement of the teaching staff and, desirably, arousal of general interest, the next task is to focus that interest (Figure 3). Any course is a "community" in which teacher and students interact with each other and with the content (Hyman, 1970); so, in an ideal situation, one would see the students of the course becoming involved at this stage. However, in practice, the staff seem to require a clearer picture of the objectives and requirements of the evaluation before involving the students. Initially, a set of objectives at the conceptual level (Scriven, 1967) are drawn up, and then a description of the course is prepared (Davis, *et al.*, 1974). The broad objectives, together with the description of the course, allow the staff and students to focus their attention on the likely tasks, and provides further framework for communication. With a large course, a working party of staff and student representatives, together with an adviser from the Centre, is established to plan the study. In undertaking this task, it is useful to consider the course as a system having

(a) inputs such as the talents, aspirations, experience and attitudes of students and teacher, as well as other socio-economic factors which have been shown by Miller (1970) to affect student attainment;

(b) operations - the conditions of learning, and the procedures and techniques used to manage this condition;

(c) outputs which refer to the outcome of the course as expressed by changes in the students' attitude to the subject as well as achievement.

---

FIGURE 3

---

The consideration of a course in terms of its inputs-operations-outputs provides an approach to the examination of the course which ensures that all factors are considered and fully investigated in a systematic manner. It also assists the reporting of the results back to the Department or lecturer in a manner which provides for solutions that take into account likely effects on other aspects of the course and possible consequences in regard to administration (Clift, 1969).

Within the framework of the system, objectives for the study are refined and developed to the manifestational level. From these objectives specific questions are then formulated, and the methods for gathering the data decided upon. It is at this stage that the boundaries of the study are delineated. For example, the study might be limited to a short period of time, some areas of the course may be seen as being outside the evaluation, or there may be restriction as to the techniques that can be used: some of the staff may not agree to the use of videotape, there may be restrictions on access to data, and some may not agree to specific results being made public. It is important to remember that, if the analysis of the requirement is inadequate, then the information gathered is likely to be insufficient and, while providing a description of the course, does not provide for the analysis in depth necessary to justify responsible decisions.

While the goals of the apprehending stage are to involve people and to focus attention, its objectives can be listed as:

(a) to inform course members of their involvement;

(b) to give course members confidence in the coverage, reliability and practicability of the study;

(c) to lay the foundation of a "climate" responsive to change;

(d) to motivate course members to move into the acquisition stage;

(e) to develop appropriate relationships in the group for open discussion and learning.

## Acquisition Stage

The acquisition stage provides for the analysis and interpretation of the collected information. The effectiveness of this stage will depend upon the quality of the data collected, the validity of the interpretations offered, and the timing and quality of the feedback to the teaching staff and students. The necessary input into this stage of an evaluation study is shown in Figure 4.

FIGURE 4

Obviously the quality of the data collected will depend upon the instruments and procedures used, but it is also important that these procedures appear appropriate to the course members, otherwise the quality of the data will be suspect (Barber, 1973). In addition, if the information is to be meaningful to the course members, it needs to be drawn from a variety of sources (Figure 5).
By the use of inventories, questionnaires, interviews, observation, etc., the validity of the results can be increased. However, the use of a variety of methods, each supportive of the other, is a self-perpetuating system and one in which a judgement is called for as to when the chain can be broken. That is when the value of the information gathered is considered less important than the resistance of the staff or students to its gathering.

FIGURE 5

The evaluation situation calls for a much greater planning and control of the data gathering than is often the case in a more experimental situation where the timing of data collection can often be manipulated to meet the demands of the research design. In an evaluation study, the stages at which information can be gathered are generally dictated by course planning, and so the most favourable times for collecting information may occur on specific days throughout the course. This implies advanced planning to ensure the efficient gathering of information with least interruption to course work and minimum demand on student and staff time. The most effective method of planning data gathering is by the use of a logical step diagram; if the project is a complex one involving printing requirements and computer programming, the critical path control method is highly suitable. By use of these control techniques, it is possible to ensure:

(a) the gathering of information at an appropriate time in relation to other course activities such as examinations, assignments, conclusions of particular phases of the course;

(b) the least demand is made on staff and students by gathering information in pre-planned sessions rather than making continuous demands on staff and students throughout the course. Too many demands for information lowers the degree of co-operation of respondents and their willingness to comment;

(c) a rapid processing of the information and, hence, an effective feedback to both staff and students. It has been found when conducting evaluation surveys that the interest and co-operation of the staff and students is increased when they receive some feedback from the information they have provided. Co-operation is further improved when the feedback results in some positive action such as lecture improvement and better inter-relationship between subject areas.

Because one is using a considerable amount of subjective comment in evaluating a course, it is often necessary to gather a certain amount of additional data during the analysis stage. If extra information is required from the students, this can often be obtained through a sample of students - in fact, if in the planning of the evaluation a sample of students is drawn and asked to participate as a representative group, then the gathering of extra data can be achieved with minimal disturbance of the student population. The process of interpreting this type of data involves the steps of identifying and separating issues and relating common elements.

By selecting issues in this way the complexity of the information is unravelled and the inter-relationship between the various issues revealed. This enables the identification of the significant issues and interactions for the evaluation. Analysis must be followed by synthesis which involves combining apparent non-related elements into a meaningful whole. In this manner, the contingencies existing between the input, operations, and output sub-systems of a course can be identified as well as the congruence between the stated course objectives and the observed behavioural changes.

It is necessary, in collecting data, to use a range of techniques and procedures. The same is true when staff and students come to consider the results. If they are not to suffer from "mental indigestion", the adviser must organise the data so that it is meaningful. McKeachie (1976) uses Katz's concept of "mental dazzle" to describe the reaction to massive computer print-outs which he sees as being detrimental to successful communication of information, inhibiting its assimilation and, hence, limiting its application in decision-making.

## Application Stage

To maximise the learning resulting from the evaluation study, it is considered necessary to go beyond a simple reporting of results. While discussion sessions are a useful adjunct in reporting the findings, the use of 'teach-ins' (Clark and Imrie, 1977) has been found to have greater effect on the degree to which the information is used. Such a teach-in not only considers the data derived from the study but provides for the sharing of experiences of the evaluation. During such a teach-in, problems can be identified, solutions proposed and discussed, and points agreed as proposals for short-term and long-term course development. If subsequently the evaluation study has been an effective learning experience and has had impact, there will be noticeable effects within the course, and those who have been engaged in the study will be able to transfer their new knowledge and skills to other courses. It is this latter component, in which knowledge and skills are applied in other situations, which suggests that evaluation studies are potentially one of the most effective methods available to institutions for in-service development of teaching competencies.

## SUMMARY

Institutions of higher education are now realising the need for continuing internal evaluation of the learning experience which is provided for their students. This awareness has resulted in the establishment of research and advisory units within tertiary institutions. A search of the literature provides numerous warnings of the ineffectiveness of educational research in providing for positive change. One of the major reasons given for this failure is the ineffective communication between the educational research worker and the educational planners. In describing the process of communication, Berlo (1960) lists a number of factors affecting the fidelity of any communication. The important factors are the attitudes of the receiver, the skill and knowledge of the communicator as well as his position within the system. These same factors are all essential components of an evaluation study. In assisting a Department with an evaluation study, a major task of the adviser is to develop a climate responsive to learning and change. To achieve such an objective, the adviser must have credibility in the eyes of his colleagues, that is they must respect his level of knowledge and skills, and his ability to provide sound practical advice. But what is sometimes overlooked, when factors affecting the communication process are discussed, is the adviser's own role/attitude in the evaluation process - objective involvement and professional commitment.

As Dressel (1961) points out, education is a complex process and, as choices have to be made, evaluation is inevitable.

> "Education is a complex process involving the selection of ideas (concepts, values, skills) and the planning of experience designed to foster mastery of these ideas in the people subject to the educational process. Choices must be made in planning an educational programme, and the effectiveness of the programme must also be studied. Evaluation is, therefore, inevitable in education."
>
> (p. 6)

And, again, Dressel (1976): "Evaluation captures the very essence of education" (p. 9).

This paper is concerned with the technology of evaluation, that is the application of systematic procedures that have clear purposes and lead to demonstrable results. It does not attempt to provide a theory to link cause and effect. The model is one that has evolved over a period of years of involvement in evaluation studies in a variety of disciplines and at a number of levels. Such a model views evaluation as a learning situation in which staff, students, and advisers learn how to handle and improve the contingencies of a course. An essential element is the quality of feedback, and is dependent on meaningful communication. Effective communication can only occur in a receptive climate, and starts with careful planning, including the definition of objectives. It is aided by the use of models which, by providing the basis of control, collaboration, and completion, can be applied in any tertiary institution and are capable of accommodating variations as well as variables.

This paper has described one such model.

## References

BARBER, T. X. (1973). "Pitfalls in Research: Nine Investigator and Experimenter Effects", in *Second Handbook of Research on Teaching*, R. M. W. Travers (Ed.). Rand McNalley & Co.

BERLO, D. K. (1960). *The Process of Communication.* Holt, Rinehart and Winston.

CLARK, M. J. *et al.* (1977). "A Strategy for Evaluating University Courses", *Occasional Paper*, 3, University Teaching and Research Centre, Victoria University of Wellington.

CLARK, M. J. and IMRIE, B. W. (1977). *Some Details and Some Consequences of an Evaluation Strategy.* University Teaching and Research Centre, Victoria University of Wellington.

CLIFT, J. C. (1969). "An Approach to Training Research", *Technical Education*, 11, 2, 60-61.

CRONBACH, L. J. (1963). "Course Improvement Through Education", *Teachers' College Record*, 64, 8, 672-683.

DAVIS, R. H., ALEXANDER, L. T. and YELON, S. L. (1974). *Learning System Design.* McGraw-Hill Book Co.

DI VESTA, F. J. (1970). *Instructional Strategies: Multivariable Studies of Psychological Processes Related to Instruction.* Pennsylvania State University.

DRESSEL, P. L. (1961). "The Essential Nature of Evaluation", in *Evaluation in Higher Education*, Dressel and Associates. Houghton Mifflin Co.

DRESSEL, P. L. (1976). *Handbook of Academic Evaluation.* Jossey-Bass Publish.

DRESSEL, P. L. and LORIMER, M. F. (1961). "Institutional Self-Evaluation", in *Evaluation in Higher Education*, P. L. Dressel and Associates, pp. 393-432. Houghton Mifflin Co.

GOLDSCHMID, M. L. (1977). "The Evaluation and Improvement of Teaching in Higher Education", Keynote address presented at the Third International Conference on Improving University Teaching, Newcastle-upon-Tyne, June 8-11.

HYMAN, R. T. (1970). *Ways of Teaching.* J. B. Lippincott Co.

KATZ, D. (1950). *Gestalt Psychology.* New York: Ronald Press.

McGEHEE, W. (1957). "Some Unanswered Questions in Training Research", *Methods of Research in Technical Training*, (Ed.), Dubois, P. H. and Manning, W. H., pp. 66-72.

McKEACHIE, W. J. (1976). "Psychology in America's Bicentennial Year", *American Psychologist*, pp. 819-833.

MILLER, A. H. (Ed.). (1977). "Symposium: Teacher Education for Tertiary Teachers", *The South Pacific Journal of Teacher Education*, 5 (1), pp. 5-63.

MILLER, G. W. (1970). *Success, Failure and Wastage in Higher Education.* G. Harrap & Co.

SCHUTZ and SILBERMAN. (1966). "Changing the Contingencies Operating on the Researcher", *Harvard Educational Review*, 36 (3), pp. 297-301.

SCRIVEN, M. (1967). "Methodology of Evaluation", in Tyler, Gagne and Scriven, *Perspectives of Curriculum Evaluation.* Rand, McNally & Co.

FIGURE 1. EXPLOSION IN PUBLICATION RELATED TO IMPROVEMENT THROUGH EVALUATION

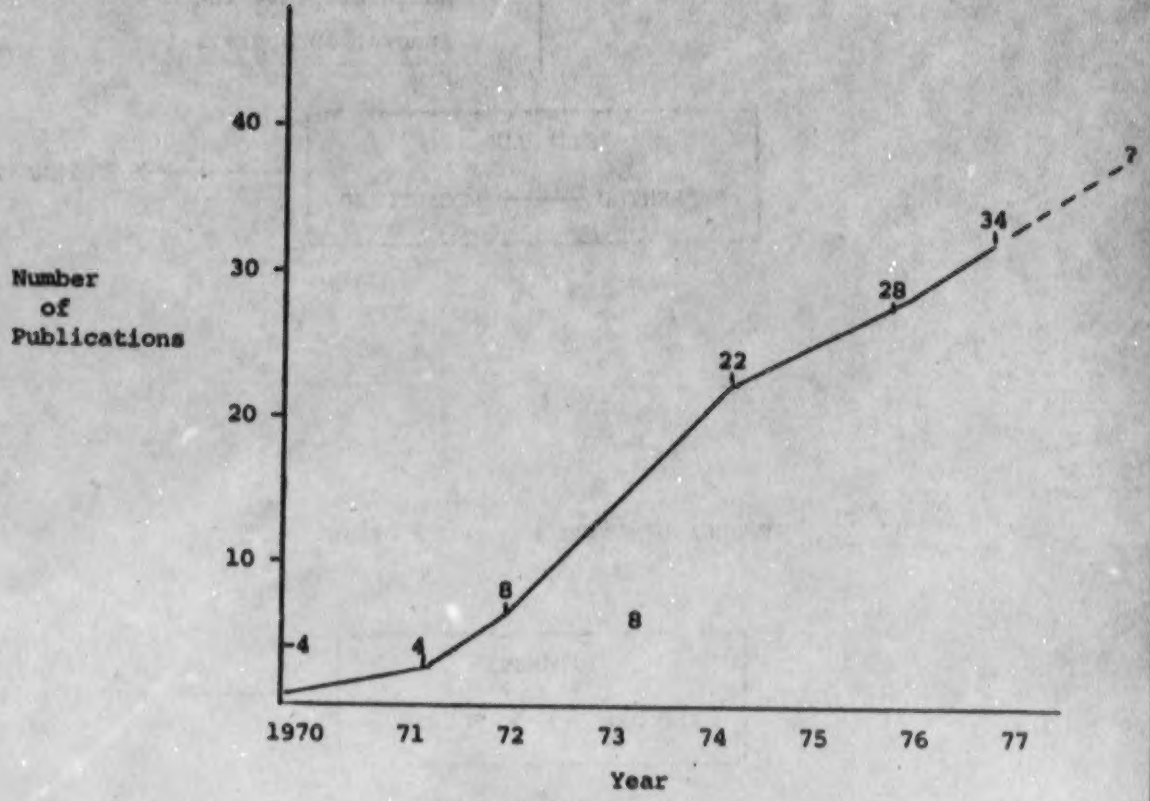


**FIGURE 2.** APPREHENDING STAGE: STIMULUS OF INTEREST

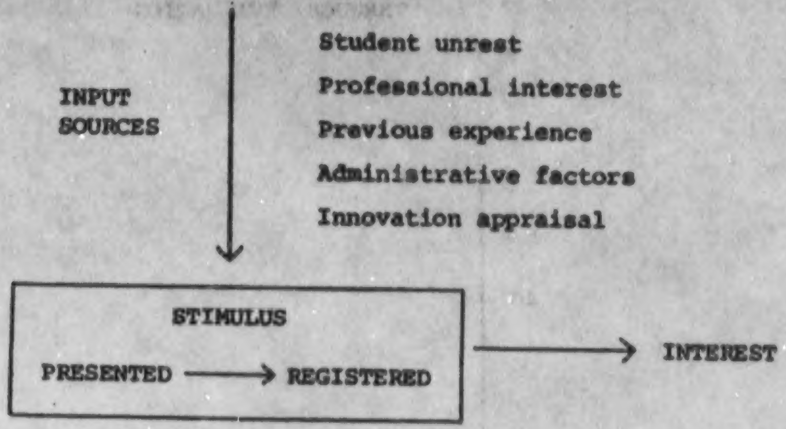

INPUT SOURCES
Student unrest
Professional interest
Previous experience
Administrative factors
Innovation appraisal
STIMULUS
PRESENTED
REGISTERED
INTEREST

**FIGURE 3.** APPREHENDING STAGE: FOCUS OF INTEREST

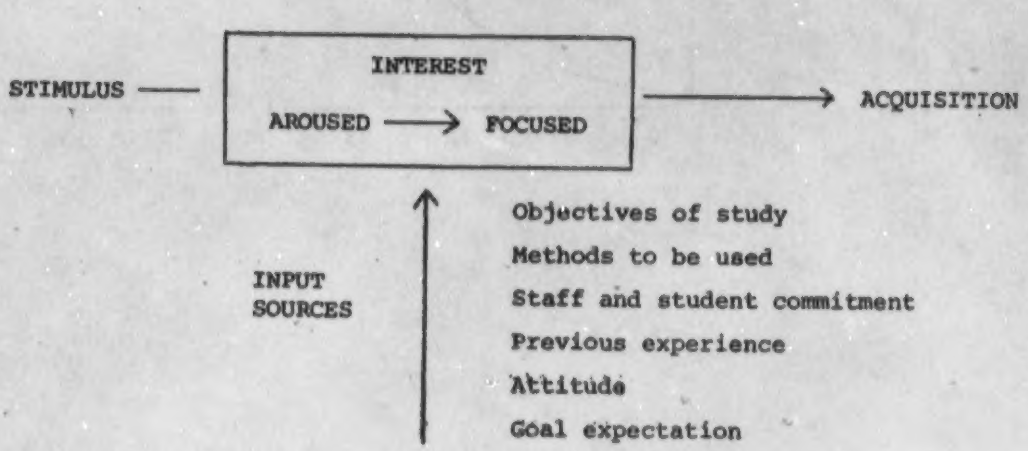

STIMULUS
INTEREST
AROUSED
FOCUSED
ACQUISITION
INPUT SOURCES
Objectives of study
Methods to be used
Staff and student commitment
Previous experience
Attitude
Goal expectation

**FIGURE 4.** SOURCES OF INPUT AFFECTING ACQUISITION STAGE

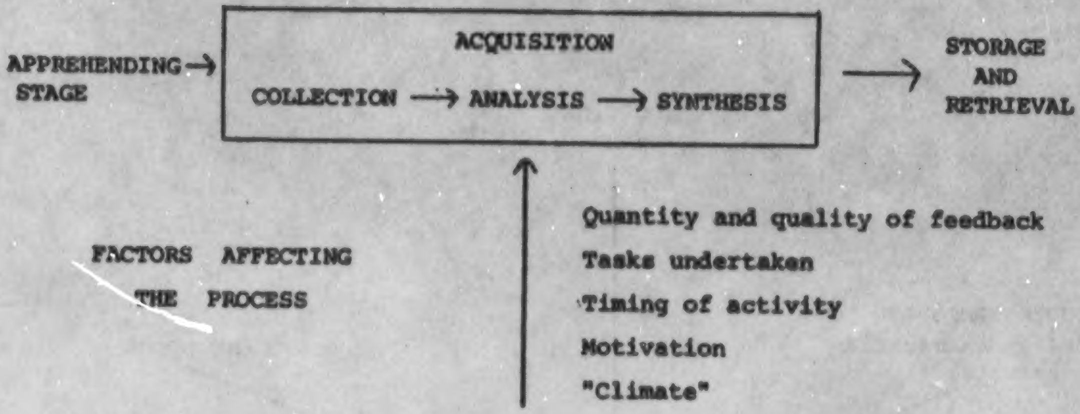
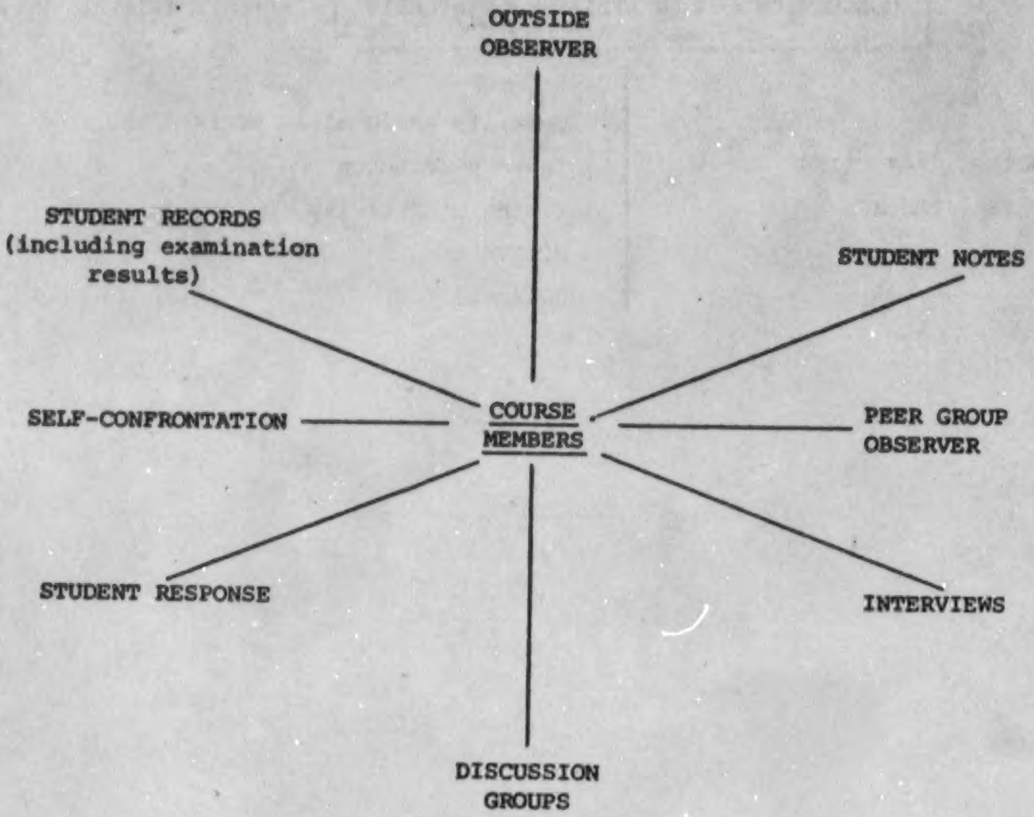

APPREHENDING STAGE
ACQUISITION
COLLECTION
ANALYSIS
SYNTHESIS
STORAGE AND RETRIEVAL
FACTORS AFFECTING THE PROCESS
Quantity and quality of feedback
Tasks undertaken
Timing of activity
Motivation
"Climate"

FIGURE 5. SOURCES OF INFORMATION


OUTSIDE OBSERVER
STUDENT RECORDS (including examination results)
STUDENT NOTES
SELF-CONFRONTATION
COURSE MEMBERS
PEER GROUP OBSERVER
STUDENT RESPONSE
INTERVIEWS
DISCUSSION GROUPS

# PUBLIC AND PRIVATE: SCHOOL SYSTEM EFFECTS ON SOCIAL AND ECONOMIC ATTAINMENTS IN AUSTRALIA*

Trevor Williams and Patrick McDonnell
Department of Sociology
Research School of Social Sciences
Australian National University

Some 21 percent of Australians receive their education in non-government schools; the majority in schools conducted by the Roman Catholic Church (17 percent), the remainder in schools set up under the auspices of various Protestant churches. All three school systems teach to essentially the same curriculum, that provided by the State Department of Education. The essential differences between the systems stem from the context in which this teaching takes place, itself the rationale for the existence of the independent schools: state schools provide a free and secular education; Roman Catholic schools exist because they believe a moral/religious component to education to be essential; and, while the Protestant independent schools adopt a similar attitude toward the moral/religious component, though perhaps with less fervor, they offer in addition the traditionally accepted virtues of the great English Public Schools.

In the investigation reported here we examine the question of whether, and how, the school system context in which formal schooling is acquired has affected the subsequent social and economic attainments of a national sample of males in the workforce in 1973: in essence, the question of whether education in an independent school pays off in enhanced status and income *vis a vis* education in a State high school and, if so, why?

## State Schools and Independent Schools

The 200 year evolution of educational institutions in Australia has been well documented (e.g., Partridge, 1968; Hansen, 1971). The first educational institutions were all "independent" schools modelled on the British Grammar School, designed to provide a classical, scientific, and religious education for sons of the colony's elite, and supported by Government funds. A concern with the need to provide education for all produced increasing numbers of secular and free Government schools and, in the 1850's, a dual education system in which both State and Independent schools were supported out of public funds. Separation of Church and State came about with the major education acts passed in the late 1800's, and had the effect of separating independant schools from Government financial support. Only within the last twenty or so years have independent schools received public funds once again.

Although the advantages that stem from education in an independent school are not well documented, they are enthusiastically endorsed (e.g., McKeown and Hone, 1967). For Roman Catholic schools the proclaimed advantages stem from the fact that they provide for religious and moral development within the context of their religious beliefs at the same time as they provide for intellectual development. The anticipated outcome is something like a "community of Catholics", individuals whose integrated intellectual/moral/religious education has produced a sense of Catholic identity (Bourke, 1976).

The noncatholic independent schools maintain the traditions of the British Grammar schools in that they are designed to provide an elite education for the offspring of the social and economic elite. The virtues that built and maintained the British Empire are the virtues developed in these schools: discipline, self-reliance, courage, judgement, integrity, adaptability -- all the virtues necessary to produce a Christian gentleman/woman, and the leaders of the nation (Slim, 1967). Integrated with this is a strong emphasis on academic excellence whose effects are evident (Hansen, 1971: 84). In short, just as Roman Catholic schools foster a "Catholic" identity, the noncatholic independent schools aim toward the development of an identity that can only be described as "elite".

Given this, it seems reasonable to postulate a number of possible mechanisms by which individuals educated in independent schools may be advantaged in the attainment of occupational and economic status. First, the noncognitive aspects of education in an independent school may be seen as *adding* to the educational resources that individuals can convert into status. In this sense, these act like "more" education. Second, the conversion of educational attainments into occupational and economic attainments may be somewhat different for the three groups in question either because the "quality" of education differs (or is seen to) between systems making a matriculation from one system worth more than the equivalent in another, or because the informal network of contacts established in the independent schools facilitates the conversion of educational attainments into status and income. Third, it is possible that the student populations of the three school systems are indicative of three identifiable subpopulation groups -- relatively wealthy Protestants, the more devout of Roman Catholics, and the rest -- in which the processes of status attainment are different overall, in the same sense that status attainment is different for men and women.

---

*Draft: please do not cite or quote without permission.

In the analyses that follow we examine the tenability of such speculations, and address the question of whether, and how, the context in which secondary education is attained affects subsequent status attainments.

## Data and Variables

We drew on data from the ANU Social Mobility in Australia Project which is based on a national sample of 4939 persons aged between 30 and 69 years (see Broom et al., 1977 for details). The present analyses are limited to the 2715 men economically active in 1973 (the year of the survey). Compensation for nonresponse and unequal probability of selection reduces this to an effective sample size of 2560.

The first substantive analyses to emerge from this project are based around a basic status attainment model that links attainments in one generation (father's education, father's occupation, mother's education, family size, and wealth) to those in the next (respondent's occupational status and income) through "years of basic education completed", "first job status", and "years of additional education", achieved career characteristics of the respondent (see Jones et al., 1977). School system effects are examined within the context of this model. The form of the equations which describe the model can be seen in Table 2.

Occupation. The respondent's first (JOB1) and current (JOBM) occupations, along with his report of his father's occupation at (respondent's) age 14 (DADJOB), are coded in terms of a three-digit occupational status score developed for Australian occupations (Broom et al., 1977: ch.5).

Income. The respondent's current income was measured on a 16-point scale which began at less than $20 per week, increased at $10 intervals to $120, by $20 intervals to $160, and by $40 intervals to $240. Dollar amounts representing the midpoint of each range were assigned to each score with the two extreme scores given average values reported in those ranges in the 1974 Current Population Survey.

Education. Father's education (DADED) and mother's education (MOMED) are treated as age at leaving school, less 6 years. The respondent's education was treated in two ways: as years of basic education (BASED), a composite of formal schooling and other education taken prior to first job; and, as levels of education completed - intermediate certificate (INT), matriculation (MAT), tertiary nonuniversity (TERT), and university (UNIV). "Years of additional education" completed after entry into the labor force (ADDED) is defined in a way similar to BASED.

School System. Respondents reported on the type of school system attended. For the present purposes we will distinguish the three systems as follows: state (ST), Roman Catholic (R ), and "Other Independent" (IND).

Family Size. Respondents reported the number of siblings they had. Respondents reporting more than eight brothers or eight sisters were coded as eight, so that maximum family size is seventeen.

Wealth. This variable is a composite index used as a proxy for parental income, and is based on questions about ownership of consumer durables.

## Background and Levels of Attainment

The effective sample size of 2560 was reduced to 1836 by restricting it to those who had completed their secondary education in Australia. Table 1 presents the means and standard deviations for the basic model variables for each of the three school-system defined subgroups and for the total group.

[Table 1 about here]

Means for the group as a whole do not differ markedly from those for all 2560 economically active males sampled (Jones et al., 1977). However, as might be expected the group under examination here is slightly better off: the average of their family possession index is about 15 percent of a standard deviation higher; their current job status is about 10 points higher; and their income is about $5 a week greater. For other variables the differences are negligible. Thus, while the analyses reported here refer strictly to the group of economically active males who completed their education in Australia, a generalization of the findings to the total economically active population of Australia (in 1973) seems not unreasonable.

Also as expected, there are differences among the means of subgroups defined by school system attended. For all but one of the variables in question catholic school means lie between state and independent school means, and, on the whole, are closer to the state school values: the one exception is for father's occupation where the catholic group is slightly (8 status points) lower than the state school group. The analogous values for independent school students exceed those of the other two groups without exception: on the average, family size is about one child lower; the average years of basic education achieved exceeds that of state school students by almost three years; differences in occupational status attainments are marked as well, particularly with regard to father's status and respondent's first job status; and finally, there is more than a $50.00 difference in average weekly income.

In brief, in comparison with those respondents educated in state and catholic schools, the products of independent schools in Australia come from more prosperous backgrounds and remain a more prosperous group. (They are also inclined to be more variable, as an examination of the standard deviations shows, particularly in job status and to some extent in income as well.)

## School System Effects: The Linear Additive Effect of School Type

In this section we consider whether attending a catholic or independent school has any additive effect, *vis a vis* attending a state school, on the four outcomes of the basic model: first job status, additional education, current job status, and income. (The equation to BASED is included as well to specify the complete model.) To do this dummy variables were constructed to represent school type: IND stands for a nominal dummy variable which assigns a score of 1 to those who have been in independent schools, and zero to others; RC is an analogous variable in which catholic school students are scored as 1, and all others zero. To avoid linear dependence, a dummy for state school attendance is omitted from the equations shown in Table 2, and effects of school type are with respect to the omitted group, state schools. Thus, the effect of IND (or RC) is interpreted as the advantage that accrues from attendance at an independent school (or catholic school) *vis a vis* attendance at a state school.

[Table 2 about here]

The results of estimating equations to JOB1, ADDED, JOBN, and INCOME including IND and RC along with the other variables specified by the basic model are shown in Table 2. Consider first, the equation to JOB1, the most direct outcome of schooling. Only father's occupation, basic education and IND have effects large enough to be considered statistically significant (each greater than twice its standard error). These are interpreted as follows: each additional 100 points of father's status adds nearly 17 points to the son's JOB1 status; each year of basic education adds about 22 points; and having been to an independent school rather than a state school contributes an additional 21 points. Thus, among individuals from similar socioeconomic backgrounds, and with equivalent years of education completed, those who attended an independent school are advantaged (*vis a vis* those educated in state schools), in terms of the status of their first job, to the extent of some 21 status points. Potential explanations for this finding are several, and obvious, and we discuss some of the possibilities below.

Additional education (ADDED) is not well predicted by this model and, for the most part, the coefficients are insignificant. In the equation predicting JOBN, neither of the school system variables (IND, RC) have statistically significant effects, though if we could assume some reliability for the IND coefficient, which is still sizeable (16.945), it would suggest that there is an advantage of independent school attendance which carries through one's career and continues to affect occupational attainments subsequent to the first, both directly and indirectly. More strictly, we are limited to the interpretation that gaining one's education in an independent school influences occupational status only on entry into the labor force. Effects on subsequent status attainments are indirect via JOB1, and somewhat attenuated as a result.

In the equation for INCOME neither school system coefficient achieves statistical significance. School system effects on economic attainments appear to be indirect only, and via occupational status attainments.

Thus, the equations estimated in Table 2 suggest the following observations about school system effects on attainments: first, education in a Roman Catholic high school offers no discernible advantage, in these respects, over education in a State school; and second, benefits in the form of enhanced first job status seem to accrue from taking one's secondary education in an independent school (*vis a vis* a State school), but are limited to this, and to the relatively minor advantages that follow from the effects that first job has on subsequent attainments.

## An Elaboration of the Measure of Education

Given what we know and/or suspect about the way in which educational attainments are converted into occupational attainments, years of basic education completed seems a fairly crude and somewhat insensitive way in which to consider educational attainment. It assumes, for instance, that each year is worth the same in terms of occupational attainment, and ignores the apparent importance of the educational certification that goes with the completion of intermediate and matriculation level examinations (or their equivalents). Moreover, the variable contains postsecondary education as well as the high school education that is of interest here, and is less than suitable as a result. Thus we may be disguising the influence of school system context as a function of the way in which we treat the measure of education.

Extended analyses aimed at the correct specification of the education variable suggest that treating it as a series of steps defined by educational credentials -- intermediate certificate, matriculation, tertiary nonuniversity, and university -- is appropriate: for example, these analyses show that the years between an intermediate certificate and the completion of a matriculation count for little more than the intermediate certificate itself in the determination of status and income. The conversion of education into status and income is the conversion of educational credentials.

In Table 3 we show the results of estimating basic model equations in which education is treated as a series of ordinal dummy variables (Lyons, 1971). The variable INT is a dummy in which all individuals with an intermediate certificate or more are coded 1 and all others zero. Again, in the case of MAT all those with a matriculation or more are coded 1 and all others zero. TERT and UNIV are treated similarly with the result that education is represented now by four variables. Coefficients associated with these variables in Table 3 may be interpreted as the effect of moving from one level of education to the next.

By treating education in this way we provide evidence on one further point as well. It seems highly plausible that the apparent influence of the elite independent schools on attainments is a function of the fact that more of their students go on to tertiary education than do the graduates of State schools. In other words, the effect of school context on occupational and economic attainments may be indirect through the attainment of higher education. The first two equations shown in Table 3 address this question.

[Table 3 about here]

The coefficients for the first two equations support the latter proposition, at least for the noncatholic independent schools. In comparison with students graduating from State high schools, those from independent schools are more likely to go on to tertiary education and more likely still to attend university. Attendance at a Roman Catholic school offers no advantages in this respect over a State school education. Beyond this in the model, effects of school system attended are negligible and, thus, within the constraints of a simple additive model, school system effects on attainments are indirect and a function of the increased likelihood that students from noncatholic independent schools will attain a tertiary level education.

Because the focus of these analyses is on school system effects on status attainments we propose to concentrate on these to the exclusion of a full interpretation of every coefficient in the model. However, an interpretation of the coefficients for the education variables is of particular interest and provides an illustration of the way these ordinal dummy variables are used to specify the effect of education. In terms of effects on the status of first job, gaining an intermediate certificate is worth some 18 status points, gaining a matriculation an additional 62 points, tertiary education is worth an extra 109 points, and a university degree 73 points on top of this. With the exception of TERT, these education variables continue to have direct effects on subsequent education (ADDED) and on occupational status (JOB1, JOBN), but only the attainment of a university degree affects income directly -- other things equal, university graduates are paid more, some $61 a week more, in fact.

## "Quality" of Schooling

In additive models of the kind estimated in Tables 2 and 3, the influence of school system context is constrained to have an additive effect over and above that due to educational attainments themselves: gaining a matriculation affects the likelihood of obtaining tertiary level qualifications and has the same effect for all students irrespective of school system attended; attendance at an independent school adds to this likelihood for the group of students from these schools.

But, the whole rationale for the existence of the two independent school systems rests on the notion that they provide an education that differs in quality from that of the secular State schools because of the context in which the learning takes place. For the purposes of these analyses we define "quality" as the marketable quality of the education obtained, the extent to which equivalent levels of formal education are differently valued and rewarded in terms of status and income because the quality of education differs between school systems, and/or is seen to by employers. At least four possible sources of differences in quality seem plausible: some school systems are simply better at teaching the traditional skills that all teach; school systems vary in the extent to which they impart nontraditional -- but eminently marketable -- skills over and above the traditional curriculum (achievement orientations, leadership qualities, independence, and the like come to mind as being associated with independent schools); school systems differ in the training, information, and guidance they provide with respect to their students' entry into the labor force; and, the select populations of the two nongovernment school systems provide informal opportunity networks that enhance the marketability of educational attainments. In short, the marketable quality of education may differ between school systems -- and/or is seen to differ by employers -- such that equivalent levels of formal education are differently valued and rewarded in terms of status and income.

With the data at hand we cannot look at these issues in detail, only at the question of whether educational attainments count more or less for subsequent occupational and economic attainments as a function of school system attended. The strategy adopted to examine this question -- the question of statistical interaction between educational attainment and school system in their effects on status and income -- is that developed in Bottenberg and Ward (1963), Kelley et al. (1969), Ward and Jennings (1973) among others. Dummy variables were created to represent the educational level by school system cross-classification such that educational attainment is now represented by eight dummy variables: three reflect the attainment of an intermediate certificate in each of the three school systems (I-ST, I-RC, I-IND); three more are the analogues of these for matriculation (M-ST, M-RC, M-IND); and the remaining two variables represent tertiary level education (TERT, UNIV) defined as before. The results of estimating equations containing these "education-by-system" terms are shown in Table 4.

[Table 4 about here]

These data suggest that the preceding additive models present an oversimplified picture of the way in which educational attainments and their context affect subsequent status attainments. The estimates shown for the equations to TERT and UNIV indicate that a matriculation taken in an independent school increases the likelihood of undertaking tertiary level education (vis a vis a matriculation from a State school), and among the independent schools, graduation from a noncatholic school produces the greatest likelihood of all. These data are not so important for the marketability argument as they are to show how school system differences in the likelihood of attaining a tertiary level education affect subsequent attainments indirectly.

In the equation to JOB1, the direct effects of having an intermediate certificate or better (vs. a lower qualification) are equivalent for state and RC groups and negligible for the noncatholic independent school group (as a function of the fact that only 8 percent of this group failed to achieve this qualification): an intermediate certificate or its equivalent produces some 19 status points, other things equal. However where a matriculation level qualification is concerned the situation is different. Among individuals of similar socioeconomic background and with equivalent levels of tertiary education, those who matriculated from the noncatholic independent schools are advantaged over those from State schools, and both are advantaged over the graduates from Roman Catholic schools, with respect to occupational status or entry into the labor force: other things equal (including tertiary qualifications), a matriculation from a noncatholic independent school is worth some 78 status points, matriculating from a State school counts for 67 points, but graduating from a Roman Catholic high school results in a gain of only 37 status points.

Between-system differences in this respect are minimised for current occupational status (JOBN) and marginally favor the graduates of Roman Catholic schools to the tune of 62 status points versus 54 and 59 for graduates of the other two school systems. Note, however, that the between system differences in effects on status at entry into the labor force (JOB1) will be transmitted to JOBN indirectly via the effect of JOB1 on JOBN, a fairly substantial effect when it is considered that, ceteris paribus, each 100 points of JOB1 status is converted into 38 points of JOBN status.

Where income is concerned, between-system differences in direct educational effects are negligible. The only direct effect of education is the $58 premium paid to university graduates. Nevertheless, the between-system educational effects continue to operate indirectly via JOB1 and JOBN, though they are attenuated as a result.

It seems then that the "quality" of education -- in the sense of the marketability of educational attainments -- does differ between school systems, favoring the graduates of noncatholic independent schools over those from State schools, and both over matriculants from Roman Catholic schools. These differences in the "quality" of education are most noticeable in the early career in their effects on initial occupational attainments. Between-system differences in direct effects on later occupational status attainments are minimised, and, in the case of income, are limited to indirect effects via occupational status.

## School System and Subpopulation Group

Although we are not really convinced, we speculated earlier that the evolution of schooling in Australia may have produced three reasonably distinct subpopulation groups -- the more devout of Roman Catholics, relatively wealthy Protestants, and the rest -- in which the processes of status attainment differ as they do for other subpopulation groups; for example, men and women, migrants and Australian born, rural and urban groups, and so on. If so, effects ascribed to school system differences may be simply reflections of more pervasive subpopulation group differences in the processes of status attainment. Such a confusion could arise, for example, if there were separate school systems for males and females and one assumed no sex differences in the processes of status attainment.

Statistically this implies that school system, the subpopulation group indicator, conditions the effects of more than education within the basic status attainment model such that, for example, the effect of JOB1 on JOBN, or JOBN on INCOME is different in the different groups. To examine this notion we relax the constraint that variables other than education have additive effects, and test for nonadditive effects throughout the model. This kind of question can be seen also as a problem in the analysis of covariance in which group membership is the covariate. Either way it comes down to a comparison of analogous partial regression slopes across groups to answer the question of whether or not subpopulation group membership makes a difference to the force with which one variable affects another. If the slopes are different most likely the social processes they represent are different in each group.

To address this problem we adopt techniques derived within the context of the general linear model and involving the use of dummy variables to represent group membership; see Bottenberg and Ward (1963), Gujarati (1970), Stolzenberg (1974) and Specht and Warren (1976) for details of method. The techniques are similar to the more familiar use of product terms in regression equations to test for interactions among continuous variables (see, for example, Digman, 1966; Ezekiel and Fox, 1967) except that, in this case, categories of group membership are coded as dummy variables, and product terms are formed between these dummies and the independent variables in a particular equation (cf. Gujarati, 1970: 20). The dependent variable of interest is then regressed on the independent variables, all but one of the dummy variables, and the product terms defined by the included dummies (see Gujarati, 1970: 22). Interpretation of the partial regression coefficients for the product terms and the dummy variables is with reference to the omitted category of group membership and takes the following form: significant coefficients for the dummy variables themselves represent additive group ("treatment") effects in the sense that knowledge of group membership alone adds to the explanation of variance in the dependent variables; and, significant coefficients for the product terms indicate that the partial slopes differ between the groups in question or, in interaction terms, that group membership conditions the size of the partial relationship under examination. Both sets of additional terms in the equation provide an increment to explained variance interpretable as a measure of the increased explanatory power of the more complicated equation over its simpler additive form.

Although this technique has seen some use and has been resurrected repeatedly over the past fifteen years, most recently by Specht and Warren (1976), it is not without its problems. Althauser (1971) notes the interpretive ambiguities entailed in the use of product terms in regression equations, as a function of multicollinearity effects on the product term coefficients. Overall, the main problem seems to be "that the size of the interaction partial regression coefficient tends (with some exceptions) to be depressed relative to the coefficients for the individual terms" (Althauser, 1971: 465) with the result that one is likely to reject (falsely) the interaction model. Thus, it seems that we are using a conservative test in this respect.

Given that we are examining interactions across the whole model, and that we have examined the education by school system interactions in detail already, in the interests of simplicity we revert to the use of "years of basic education" completed as the measure of education. Furthermore, since the sample is limited to only 151 noncatholic independent school alumni and 249 individuals educated in a Roman Catholic high school, there seems little point to providing estimates for the basic model separately for each group as all but the largest of coefficients would not be reliably different from zero. Instead, Table 5 simply indicates whether the coefficients in the basic status attainment model differ between the three school system defined groups. All comparisons are with respect to the State school group.

[Table 5 about here]

Group differences in effects are most pronounced in the effects of the socioeconomic background variables on BASED. For the noncatholic independent school group the effects of all except DADED differ from those of the State school group. In the catholic school group differences (vis a vis the State school group) are limited to the effects of DADJOB and MOMED. The intercept for the noncatholic independent school group differs as well indicating an additive effect of school system as well for this group.

For the remaining effects within the model in all but one instance it is the noncatholic independent school group which differs from the majority State school group: in the effect of basic education on JOB1 which was noted in the preceding analyses; in the effect of family wealth on additional education completed; and in the effect that first job status has on current income. Overall then, the data seem to indicate that the social processes that convert the ascribed socioeconomic background characteristics of individuals into educational attainments operate differently in these three groups, but after entry into the labor force between-group differences in the processes of status attainment become less marked and are limited to the graduates of noncatholic independent schools. In short, there is evidence that in these school system defined groups the processes implicated in status attainment are somewhat different, but the evidence is less than overwhelming and one could not argue very convincingly that these have the status of subpopulation groups. Moreover, since most of the group differences shown here center on the causes and effects of education, the interpretations advanced in connection with the preceding analyses seem a more plausible explanation of school system effects than to suggest that the status attainment processes in these three groups are so different overall that we should begin to think of them as distinct subpopulation groups.

## School System Effects?

It comes as no surprise to find that the populations entering each of the three school systems are different in terms of the ascribed resources conferred by their social origins (Table 1). The independent schools charge fees for the schooling they provide, and between-system differences in the means of socioeconomic background variables are of the order of magnitude of average fee differences between the three systems. Furthermore, in the face of this socioeconomic selection by school system, and because this carries with it some selection on scholastic ability as well, it is not surprising to find parallel between-system differences in resources achieved during the process of status attainment. However, it is the extent to which the latter differences are a product of between-system differences in education, rather than socioeconomic selection, that is the focus of the analyses reported here.

Using a simple additive model (Table 2) we find that, over and above years of basic education completed, taking one's high school education in a noncatholic independent school is worth some 21 additional status points on entry into the labor force (vis a vis attending a State high school). Schooling in a Roman Catholic high school seems to offer no additional advantages. Furthermore, school system effects seem to be limited to this one instance and have no effect on later career attainments. A more complex additive model (Table 3) supports this same view, and suggests that at least part of the school system effect is due to the fact that noncatholic independent school graduates are more likely to undertake tertiary education and enhance their initial occupational status as a result. Again, we find no school system effects on subsequent attainments.

But, these analyses were governed by the assumption that formal education in all three systems was equivalent in terms of the status and income attainments it produced. School system effects were limited to additional, and unspecified, effects deriving from school system context -- other capabilities, informal networks, or whatever -- that made an additional contribution to status attainments.

However, parents pay the fees of independent schools at least in part because they believe that they are buying a better quality education for their children. Whether they get value for their money has been largely a matter of assumption. We were able to investigate this assumption insofar as one aspect of educational "quality" is concerned, namely the "marketable quality" of education, the question of whether equivalent levels of formal education are differently rewarded in the labor market across the three school systems.

The answer to this question seems to be that, in fact, equivalent levels of formal schooling are differently rewarded, certainly in the early career in the form of enhanced first job status for the graduates of the noncatholic independent schools. And, in this respect, the graduates of Roman Catholic high schools seem to fare worst of all (Table 4). Between-system differences in the effects of formal education were less pronounced in later career where current occupational status was concerned, and were negligible for income. We can do no more than speculate on the reasons for the relative advantage of noncatholic independent school graduates, and the following explanations are suggested: quite simply, their fees buy better teachers, smaller classes and better facilities and, as a result a better education at the same formal level; possibly, the education is simply seen as better by employers (or both is and is seen); the noncognitive teaching/learning so emphasized in the independent schools is the marketable commodity that everyone thinks it is; and/or the conversion of educational into occupational attainments is facilitated by the informal networks established in these socioeconomically select schools. To some extent these school system differences may be confounded by between-system differences in average scholastic ability but, given the modest direct effects of ability on attainments net of education, this bias is likely to be insubstantial.

In short, we seem to have found empirical support for what everyone already "knew": it is possible to buy a better education, in the sense of a more marketable education (possibly in other aspects as well), but it seems only marginally better in terms of the social and economic attainments it produces, and these are limited to the early career.

## References

Althauser, Robert P.

1971 "Multicollinearity and non-additive regression models." Pp.453-472 in Hubert M. Blalock Jr. (ed.), Causal Models in the Social Sciences. Chicago: Aldine Atherton

Bottenberg, Robert A. and Joe H. Ward Jr.

1963 Applied Multiple Linear Regression. Report PRL-TDR-63-6 Air Force System Command, Lackland Air Force Base, Texas.

Bourke, J.

1976 "Catholic Schools in Australia." Pp.27-40 in R.E. Browne and D.J. Magin (eds.), Sociology of Education. Melbourne: MacMillan.

Broom, Leonard, Paul Duncan-Jones, F. Lancaster Jones and Patrick McDonnell

1977 Investigating Social Mobility. Canberra: ANU Press.

Digman, J.M.

1966 "Interaction and non-linearity in multivariate experiments." Pp. in R.B. Cattell (ed.), Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally.

Ezekiel, M. and K.A. Fox

1967 Methods of Correlation and Regression Analysis. N.Y.: Wiley.

Gujarati, Damodar

1970 "Use of dummy variables in testing for equality between sets of coefficients in linear regressions: a generalisation." The American Statistician 24 (December): 18-22.

Hansen, I.V.

1971 Nor Free Nor Secular. Melbourne: Oxford University Press.

Jones, F. Lancaster, Patrick McDonnell, and Trevor Williams

1977 "Occupational Achievement and Economic Rewards: Some Social Determinants." Pp.53-75 in K.A. Tucker (ed.), Economics of the Australian Service Sector. London: Croom Helm.

Kelly, Francis J., Donald L. Beggs and Keith A. McNeil

1969 Multiple Regression Approach. Carbondale, Illinois: Southern Illinois University Press.

McKeown, P.J. and B.W. Hone

1967 The Independent School. Melbourne: Oxford University Press.

Partridge, P.H.

1968 Society, Schools and Progress in Australia. London: Pergamon Press.

Slim, Sir William

1967 "The Education of Australians." Pp.ix-xv in P.J. McKeown and B.W. Hone (eds) The Independent School. Melbourne: Oxford University Press.

Specht, David A. and Richard D. Warren

1976 "Comparing causal models." Pp.46-82 in David R. Heise (ed.), Sociological Methodology 1976. San Francisco: Jossey-Bass.

Ward, Joe H. Jr. and Earl Jennings

1973 Introduction to Linear Models. Englewood Cliffs, N.J.: Prentice-Hall.

Table 1: Means and standard deviations for total group and for each school system defined subgroup

| | Total Group | | State Schools | | Roman Catholic Schools | | Noncatholic Independent Schools | |
|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| DADED | 8.34 | 2.64 | 8.11 | 2.34 | 8.35 | 2.57 | 10.50 | 4.09 |
| DADJOB | 510.76 | 109.30 | 502.68 | 102.22 | 493.03 | 101.70 | 621.41 | 127.83 |
| MOMED | 8.21 | 1.79 | 8.06 | 1.71 | 8.25 | 1.67 | 9.47 | 2.21 |
| WLTH | .14 | 1.02 | .04 | .96 | .09 | 1.07 | 1.19 | .88 |
| FAMSZ | 4.51 | 2.54 | 4.61 | 2.57 | 4.81 | 2.54 | 3.12 | 1.59 |
| BASED | 9.54 | 2.08 | 9.26 | 1.83 | 9.71 | 1.99 | 12.00 | 2.75 |
| JOB1 | 486.46 | 91.23 | 477.72 | 79.09 | 482.40 | 80.94 | 577.08 | 149.37 |
| ADDED | .55 | .94 | .54 | .93 | .55 | .91 | .70 | 1.09 |
| JOBN | 536.86 | 119.17 | 526.25 | 112.49 | 528.96 | 105.48 | 651.15 | 141.60 |
| INC | 133.69 | 73.89 | 127.88 | 69.60 | 134.16 | 68.72 | 188.31 | 96.63 |
| N = | 1836 | | 1437<br>(78%) | | 249<br>(14%) | | 151<br>(8%) | |

Table 2: The basic status attainment model extended to include school system variables

| | DADED | DADJOB | MOMED | WLTH | FAMSZ | BASED | JOB1 | ADDED | JOBN | IND | RC | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric Coefficients | | | | | | | | | | | | |
| BASED | .129 | ---* | .168 | .483 | -.170 | | | | | | | .261 |
| JOB1 | --- | .168 | --- | --- | --- | 21.988 | | | | 21.042 | --- | .361 |
| ADDED | .035 | --- | --- | --- | -.045 | --- | --- | | | --- | --- | .031 |
| JOBN | --- | .120 | --- | --- | -3.040 | 14.351 | .407 | 22.687 | | --- | --- | .411 |
| INC | --- | --- | --- | 6.085 | --- | 3.053 | .091 | 3.502 | .228 | --- | --- | .297 |
| Standardized Coefficients | | | | | | | | | | | | |
| BASED | .164 | --- | .144 | .236 | -.207 | | | | | | | |
| JOB1 | --- | .201 | --- | --- | --- | .501 | | | | .063 | --- | |
| ADDED | .099 | --- | --- | --- | -.122 | --- | --- | | | --- | --- | |
| JOBN | --- | .110 | --- | --- | -.064 | .250 | .311 | .179 | | --- | --- | |
| INC | --- | --- | --- | .083 | --- | .086 | .113 | .045 | .368 | --- | --- | |

*Coefficient not significantly different from zero.

Table 3: Elaboration of the education variable in the basic model with school system variables

| | DADED | DADJOB | MOMED | WLTH | FAMSZ | INT | MRT | TERT | UNIV | IND | RC | JOB1 | ADDED | JOBN | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Metric Coefficients | | | | | | | | |
| TERT | ---* | --- | --- | --- | --- | --- | .216 | | | .078 | --- | | | | .217 |
| UNIV | --- | --- | --- | --- | --- | --- | .138 | | | .114 | --- | | | | .180 |
| JOB1 | --- | .152 | --- | 4.633 | --- | 18.225 | 62.485 | 109.298 | 73.536 | --- | --- | | | | .430 |
| ADDED | .028 | --- | --- | -.074 | -.032 | .197 | .333 | --- | --- | --- | --- | --- | | | .067 |
| JOBN | --- | .122 | --- | --- | -3.102 | 28.430 | 57.306 | --- | 70.398 | --- | --- | .377 | 19.834 | | .421 |
| INC | --- | --- | --- | 6.475 | --- | --- | --- | --- | 61.304 | --- | --- | .084 | --- | .224 | .302 |
| | | | | | | | Standardised Coefficients | | | | | | | | |
| TERT | --- | --- | --- | --- | --- | --- | .412 | | | .109 | --- | | | | |
| UNIV | --- | --- | --- | --- | --- | --- | .310 | | | .180 | --- | | | | |
| JOB1 | --- | .182 | --- | .052 | --- | .099 | .258 | .237 | .136 | --- | --- | | | | |
| ADDED | .078 | --- | --- | -.080 | -.086 | .104 | .133 | --- | --- | --- | --- | | | | |
| JOBN | --- | .122 | --- | --- | -.066 | .118 | .181 | --- | .099 | --- | --- | .289 | .157 | | |
| INC | --- | --- | --- | .089 | --- | --- | --- | --- | .140 | --- | --- | .104 | | .362 | |

[*] Coefficient not significantly different from zero.

Table 4: School system effects on relationships within the basic model

| | DADED | DADJOB | MOMED | WLTH | FAMSZ | I-ST | I-RC | I-IND | M-St | M-RC | M-IND | TERT | UNIV | JOB1 | ADDED | JOBN | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Metric Coefficients | | | | | | | | |
| TERT | ---* | --- | --- | --- | --- | --- | --- | --- | .182 | .228 | .335 | | | | | | .226 |
| UNIV | --- | --- | --- | --- | --- | --- | --- | --- | .097 | .124 | .317 | | | | | | .206 |
| JOB1 | --- | .153 | --- | 4.954 | --- | 18.769 | 18.749 | --- | 66.626 | 36.861 | 77.815 | 111.283 | 73.723 | | | | .432 |
| ADDED | .028 | --- | --- | -.074 | -0.032 | .209 | --- | --- | .323 | .470 | --- | --- | --- | --- | | | .067 |
| JOBN | --- | .123 | --- | --- | -3.055 | 30.937 | --- | 35.795 | 58.677 | 61.563 | 53.548 | --- | 72.565 | .377 | 19.788 | | .422 |
| INC | --- | --- | --- | 6.553 | --- | --- | --- | --- | --- | --- | --- | --- | 58.442 | .081 | --- | .224 | .303 |
| | | | | | | | | | Standardised Coefficients | | | | | | | | |
| TERT | --- | --- | --- | --- | --- | --- | --- | --- | .267 | .191 | .375 | | | | | | |
| UNIV | --- | --- | --- | --- | --- | --- | --- | --- | .167 | .121 | .415 | | | | | | |
| JOB1 | --- | .183 | --- | .055 | --- | .101 | .057 | --- | .212 | .067 | .188 | .241 | .130 | | | | |
| ADDED | .079 | --- | --- | -.080 | -.086 | .109 | --- | --- | .099 | .082 | --- | --- | --- | --- | | | |
| JOBN | --- | .113 | --- | --- | -.065 | .127 | --- | .079 | .143 | .085 | .099 | --- | .102 | .288 | .156 | | |
| INC | --- | --- | --- | .090 | --- | --- | --- | --- | --- | --- | --- | --- | .133 | .100 | --- | .361 | |

[*] Coefficient not significantly different from zero.

Table 5: Tests for homogeneity of slopes across school system defined groups

| Dependent Variable | Independent Variables | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | DADED | DADJOB | MOMED | WLTH | FAMSZ | BASED | JOB1 | ADDED | JOBN | INTERCEPT |
| BASED | --- | CI | CI | I | I | | | | | I |
| JOB1 | --- | --- | --- | --- | --- | I | | | | I |
| ADDED | C | --- | --- | I | --- | --- | --- | | | I |
| JOBN | --- | --- | --- | --- | --- | --- | --- | --- | | |
| INC | --- | --- | --- | --- | --- | --- | I | --- | --- | |

Notes:

I = partial slope in noncatholic independent school group significantly different from State school group.

C = partial slope in Roman Catholic school group significantly different from State school group.

# RESEARCH OF THE EVALUATION PROCESS: CURRENT STATUS AND FUTURE DIRECTIONS[1]

Ralph G. Straton
The University of Sydney

Program evaluation studies have now become an important feature of the policy development and decision-making processes employed in a wide variety of areas of social concern. Important secondary sources (Glass, 1976; Guttentag and Struening, 1975) have reflected this variety by recently publishing reports of evaluation studies taken from the fields of education, health, welfare and social services, and crime and justice. Evaluation studies also take place in a wide variety of settings, focus on many different phenomena, employ diverse methodologies and are concerned with the welfare of persons who differ from each other in important ways. Never-the-less, the process followed in the conduct of evaluation studies is remarkably similar across different studies. It is a process which involves first, deciding what information to gather, next, gathering the information and finally, making the information available to some person or persons for their use.

Stufflebeam et al. (1971) have captured this three stage process in their definition of evaluation and have characterized the stages as being concerned with delineating, obtaining and providing information for decision-making. However, the same process can be detected in each of the diverse frameworks which have been proposed to guide the conduct of evaluation studies (see Straton, 1977; Taylor, 1976; Worthen and Sanders, 1973). Stake (1967), for instance, has suggested that the intents of a program be specified and used in determining what observations to make to describe the antecedents, transactions and outcomes of the program and whether the intents have been realised. He proposes that all this information be made available to program audiences for their use in making judgments and decisions. The illuminative evaluation framework of Parlett and Hamilton (1976), which emphasises the methodology of social anthropology, also displays the same process. They suggest that initial observations of a general nature, plus some anticipation of the information likely to be of interest to program audiences, should guide an evaluation consultant's decisions about what aspects of the program to focus upon both in information gathering and in the provision of information to the wide variety of audiences to whom they believe it should be conveyed. As a final example, the transactional evaluation framework proposed by Rippey (1973) suggests that there should be a continuous recycling through the delineating, obtaining and providing stages as the evaluation consultant strives to respond to the developing and changing information needs of program personnel.

This common conception of the process of conducting evaluation studies has been well articulated by Stufflebeam et al. (1971) and, in less detail, by Ten Brink (1974) who has focussed primarily upon student evaluation. In essence, each of these models of the evaluation process identifies a number of critical decisions which must be made in the planning and conduct of an evaluation study. In addition, these and other authors offer advice to assist evaluation consultants in making these decisions. However, the advice which is available is typically of a speculative kind being derived from logical analysis, informally organised experience or subjective judgment rather than being based upon the results of well grounded empirical studies (e.g. see Moursund, 1973; Weiss, 1972a).

Much of the published advice on how to conduct evaluation studies has been drawn directly from prescriptions for the conduct of research studies (Airasian, 1974; Nunnally, 1975; Popham, 1975). Although these prescriptions provide good benchmarks for evaluation studies they fail to take account of the important differences between evaluation studies and research studies. In essence, most of these differences are due to the external constraints placed upon evaluation studies, constraints which tend to be either absent from or less severe in the case of research studies. They include constraints upon time for the study, accessibility of subjects, replicability of treatments, financial resources, availability of persons with relevant expertise to assist in the study and a variety of other constraints including those imposed by the political climate within which an evaluation study is being conducted. These constraints force evaluation consultants into making decisions about the conduct of an evaluation study which represent tradeoffs between various aspects of the study. Thus, advice which indicates how a study should ideally be conducted is not sufficient. What is needed is information which would help one to estimate a priori what would be gained and what would be lost if a study or some aspect of it were to be conducted in one way rather than in another.

Take, for example, a situation where an evaluation process decision is to be made regarding how to obtain information on parents' opinions about a new curriculum. In a research study the preferred data gathering method would, in general, be the one which was expected to yield the most valid data. However, in an evaluation study one would need to consider, for each possible method, the anticipated reliability and validity of the data in relation to the degree of precision required by the purposes of the study, the cost per unit observation, the time and level of expertise required for instrument development, data gathering and data processing, the anticipated response rate and a variety of other factors. Obtaining parents' opinions via interviews, for instance, may be thought to yield data of greater validity but lower reliability and to require higher cost, more time and more highly trained assistants compared with questionnaires for which the anticipated response rate may be lower. Thus, the final choice of method would represent the best compromise between these and other considerations in the light of the overall purposes, resources and constraints of the evaluation study. But such a compromise is not likely unless an evaluation consultant can obtain information on the factors critical for his or her choice and on their inter-relationships.

For evaluation studies to be conducted in an optimal way evaluation consultants require a great deal of information relevant to the many evaluation process decisions which must be made. To be maximally useful this information should be derived from empirical studies of the evaluation process itself, i.e. it should be based upon demonstrated effects not speculation. Speculative or armchair advice has its place, particularly in suggesting lines for research of the evaluation process and for contributing to evaluation process decisions in the absence of empirically derived information. However, hopes for improving the quality and utility of evaluation studies in the long term must rest upon there being available to evaluation consultants empirical research based information relevant to evaluation process decisions.

The major concern of this paper is with empirical research of the evaluation process. It has three main purposes. First, to present a model of the evaluation process which indicates the critical areas of decision which occur in most evaluation studies irrespective of their focus, setting, methodology, guiding framework or area of social concern. Second, to use this model as a basis for classifying and reviewing some examples of empirical research studies which provide information potentially of use to evaluation consultants in making evaluation process decisions. Third, to suggest areas where evaluation process research is needed and to indicate some difficulties in the conduct of such research. When one considers the potential impact of evaluation study information on a wide range of educational and social programs affecting vast numbers of people it becomes apparent that careful consideration should be given to ways of improving the quality of evaluation studies including the conduct of evaluation process research.

# A MODEL OF THE EVALUATION PROCESS

In any evaluation study many decisions must be made concerning the way in which the study is to be conducted. Each decision will have a central purpose or concern related to a particular decision area such as the specification of the audiences of the evaluation study and their information needs or the determination of how the information is to be documented once it has been obtained. It is these decision areas, embracing many related evaluation process decisions, which comprise the basic elements of the evaluation process model (EPM) presented here (see Figure 1). The model has been developed from a model suggested by Ten Brink (1974) and is also closely related to the very detailed model of Stufflebeam et al. (1971). It should be noted that the characteristic style or strategy of an evaluation study will depend upon what alternatives are chosen at each of the decision points specified in the EPM. These choices will also determine, in a very real way, what information an evaluation study will yield and the likely patterns of utilization of this information.

It is assumed in the EPM that the process decisions made in an evaluation study are interdependent. In other words, when a particular alternative is chosen in one decision situation many other decisions will be affected both within and across decision areas. For example, the fact that secondary school students are specified as an audience of an evaluation study should have implications for decisions concerned with how the information gathered should be documented. A further assumption of the EPM is that many of the evaluation process decisions and the decision areas should be considered to be hierarchically ordered in the sense that certain decisions or classes of decision should be made before other decisions. Referring back to the example just given it will be clear that there are compelling reasons for specifying evaluation study audiences before decisions are made concerning the mode of documentation to be used. In Figure 1 the arrows indicate the assumed hierarchical relationships between the decision areas specified in the model.

## The Delineating Process

The first stage of an evaluation study involves determining what information is required to meet the purposes of the study as seen by its audiences. Of course, the purposes and audiences of an evaluation study can rarely be taken as given by the evaluation study sponsor. There will almost always be many possible audiences which should be considered, each with unique concerns which would tend to modify the purposes of the evaluation study, as Straton (1976) has pointed out. In addition, not all audiences can be considered to be of equal importance and the evaluation process will in effect place them in some priority ordering. Correspondingly, the purposes or concerns of each audience will also be put in order of priority either consciously or by default.

What the delineation process comprises at this stage is an iterative procedure whereby the audiences and the purposes which an evaluation study is to serve will be determined. As potential audiences are identified and consulted about their concerns related to the program study purposes will become apparent. At the same time, as possible study purposes emerge prospective study audiences will tend to suggest themselves. By moving back and forth between the two both evaluation study audiences and purposes should be clarified. What is to be evaluated should also become clarified during this process as those aspects of a program which are of primary concern to key audiences become apparent.

The delineation process will be complete when the information to be obtained in an evaluation study, or at least in a single delineating, obtaining and providing cycle of the study, has been specified. In order to do this Stufflebeam et al. (1971) have suggested that the decisions and judgments of concern to study audiences will need to be analysed so that information relevant to distinguishing between possible alternative courses of action or judgments can be identified. These items of relevant information will then need to be put in order of priority with respect to their perceived value or importance from the viewpoint of the audiences.

## The Obtaining Process

The primary input to the information obtaining stage of an evaluation study is the priority ordering of relevant information which was produced at the delineating stage. The output of this second stage is basic data about the program of interest, appropriately analysed. In order to obtain this data the initial step involves the determination of what information is already available and what are its characteristics. Of major concern here are questions of data accessibility, completeness, reliability, validity, cost of recovery, and comparability across different sites, sub-groups and times of being obtained and recorded.

The second step in the information obtaining process, as portrayed by the EPM, is concerned with deciding questions of when, how and from whom to obtain what information. It is, in essence, the process of designing the evaluation study. Alternative answers to these design questions, which taken in combination constitute alternative designs, are appraised at this step by considering both the constraints applying to the study and the expected information yield of each alternative design.

At the next step decisions are made regarding how to operationalise the variables specified in the study design. While the primary consideration here will be what information is required there typically will be several alternative data gathering instruments or procedures which might be used. Decisions regarding broad questions such as whether the instruments should allow norm-, criterion-, domain- or self-referenced interpretations to be made, or whether individual or group estimates are required already will have been taken in determining the study design but many more specifically measurement questions will remain. One major decision will revolve around whether already available instruments and procedures might be used or whether it will be necessary to construct them especially for the study, and if so what data collection methods should be used.

The heart of any evaluation study is the actual data gathering itself. It concerns the process of actually obtaining the samples, administering the instruments, making the observations and recording the basic data specified in the study design. In a very real sense the evaluation process decisions to be made here are practical decisions. They will be concerned with such things as the timing of the data gathering, what to do if a respondent is not available, say through illness, when data gathering is to take place, how and how often to follow-up non-responders to mailed questionnaires, and the selection and training of interviewers.

The final step in the information obtaining process is concerned with data analysis. Data analysis decisions will depend not only upon the data gathering methods used and the nature of the groups from which data was gathered, but also upon the purposes of the study and the audiences to be served. Restrictions on the appropriate methods of analysis for various types of data, e.g. ratings, rankings or classification data, are well known as are the alternative procedures which may be followed in analysing data from various types of groups, e.g. random samples, ad hoc samples or populations. However, the choice of particular data analysis methods will imply that the forthcoming indices are both appropriate to the assessment of the various decision and judgment alternatives under consideration and meaningful, or at least potentially so, to the study audiences.

## The Providing Process

This stage of the evaluation process is concerned with what Stufflebeam et al. (1971) call an interface between the evaluation consultants and the audiences of the evaluation study. The emphasis is upon communicating to the study audiences the information which the obtaining stage has yielded. The basic elements of the providing process are decisions related to the documentation and dissemination of this information. Decisions related to documentation will, among other things, focus upon the media to be used, the organisation of the information, the level of difficulty of the message, e.g. its readability in the case of written material, the degree of detail of the message, and the extent to which the design and procedures of the evaluation study are to be described.

Dissemination decisions are concerned with how the documented information will be delivered to the audiences. They will, in effect, compare alternative dissemination modes in terms of their anticipated audience coverage, credibility and acceptability to audiences, cost and potential for gaining audience attention among other factors.

### Utilisation of Evaluation Study Information

Once the dissemination activities are complete the evaluation process is also complete from the viewpoint of the evaluation consultant, unless, of course, the consultant is also an audience member. The act of evaluating, of actually making judgments and decisions, is the responsibility and the right of the study audiences who may or may not utilize the information provided through the evaluation study in their deliberations. Thus, information utilization is not a part of the evaluation process in the sense being used here, yet it is an area of very great concern in making evaluation process decisions. These decisions, like any other decision, will be made so as to optimize some criterion and the final criterion must be the extent to which audience judgments and decisions related to the program have been facilitated by the conduct of the evaluation study. In other words, the audiences must have made better decisions and judgments in the sense of behaving "consistently with some a priori set of requirements or rules to which they want their choices to conform" (Becker and McClintock, 1967, p. 239).

In a very real sense the evaluation process model presented here is concerned with decision-making. It describes the areas of decision which are of concern in the conduct of most if not all evaluation studies. Many of the evaluation process decisions will not be made formally or even consciously, and yet the way in which an evaluation study is conducted will imply that certain alternatives have been chosen in preference to others. The critical question, then, is how can these evaluation process decisions be made more optimal. Highlighting what these decisions are is an important step, and this is what many of the suggested frameworks for evaluation studies do. However, it is also essential to provide evaluation consultants with information about the alternative courses of action which might be taken at each decision point in the evaluation process. This is the function of evaluation process research.

## RESEARCH OF THE EVALUATION PROCESS

The evaluation process model presented above provides a convenient means whereby the sorts of questions upon which evaluation process research might focus can be determined and classified. It also allows for the classification of empirical research already completed which is relevant to evaluation process decisions and for the identification of high priority areas for further research. Figure 2 provides a classification scheme based upon the EPM and suggests foci for evaluation process research in each area. The first three major categories are concerned with the delineating, obtaining and providing processes and are derived directly from the EPM. The fourth category focusses upon research relating to the utilization of evaluation study information. Its concern is with how utilization may be affected if evaluation studies are conducted in various ways. The final category is concerned with research of the efficacy and utility of frameworks and guidelines which might be used to aid the conduct of an evaluation study.

## Research of the Delineating Process

Research of the information delineating process should throw light on alternative ways of specifying the information to be obtained and provided in an evaluation study. The results of such research should assist an evaluation consultant in determining the audiences and purposes to be served by an evaluation study as well as the information required for doing so. Investigations might focus upon alternative ways of identifying potential audiences and of determining a priority ordering among them, the impact upon information utilization of having multiple audiences of various types, and the degree of interest of various audiences in particular types of programs. Studies in this area are rare but Groundwater-Smith (1975), in an evaluation study of a two-teacher open-plan classroom, asked twenty members of each of six potential audiences to indicate the relative importance to themselves of each of the types of information which might be gathered in the study. Questionnaire return rates ranged from 10% to 100% of an audience sample, providing some indication of which audiences had an interest in the particular program being evaluated and, possibly, in programs of this type.

There are also a variety of questions related to the purposes of an evaluation study which might be the object of empirical investigation. Included here would be studies of methods for discovering the judgments and decisions of concern to evaluation study audiences and the alternative courses of action open to them. In addition, there might be studies of ways for determining the relative importance of these judgments and decisions, the influence which consulting with study audiences about study purposes has on information utilization, and the range of purposes identified in studies of other similar programs. Again, few empirical studies in this area have been located. One exception is a study by Spivac et al. (1975) which indicated that audiences drawn from within and from outside community mental health centres differed in the priorities they placed upon the various possible purposes of evaluation studies of a centre. A majority of both groups placed a high value upon accountability from a consumer standpoint and a relatively low value upon centre management issues and cost or equity of service delivery. However, centre staff were more interested in efficiency and effectiveness compared with the outsiders who were more concerned with community involvement.

In the area of defining the information which an evaluation study is to obtain and provide to study audiences there are many possible research questions. Research might focus upon the characteristics of alternative methods first for specifying what information is relevant to the decisions and judgments faced by study audiences, second for placing this information in order of priority, and third for choosing the information actually to be gathered. Research might also investigate the types of information which typical audiences value in evaluating certain classes of program. Two recent studies have focussed upon this latter area. Mori (1975) found that for evaluating educational materials for exceptional children teachers gave highest priority to information related to student characteristics, instructional material characteristics, and the actual outcomes of learning produced by utilization of the instructional materials with students. In the second study Nevo (1974) found that high school students, teachers and principals each rated information on program goals to be at least as important as program outcomes while information on program design and implementation was rated low overall except that students rated highly the formative use of information on program design.

## Research of the Obtaining Process

The EPM specifies four steps at the information obtaining stage of an evaluation study. These steps focus upon the decisions involved in preparing an inventory of already available relevant information, developing a design for the evaluation study, choosing or constructing the necessary data gathering instruments and procedures, and completing the data analyses. Evaluation process research related to inventory preparation would include the development and investigation of procedures for searching for relevant existing records, studies of the characteristics of typical institutional records, and the development of techniques for the appraisal of records including ways to determine their reliability and validity. To date, no empirical studies of this aspect of the evaluation process have been located despite the fact that evaluation studies from a variety of areas have used data from existing records including Smith's (1976) study of a mental health centre and Kelling et al.'s (1976) study of an innovative crime prevention program. In addition, Campbell (1976) has drawn attention to some of the potential deficiencies of this type of data.

The design of evaluation studies has been extensively dealt with in the literature (Airasian, 1974; Campbell, 1975; Scriven, 1974) and a variety of checklists, guidelines and standards have been proposed for use in developing and in judging evaluation study designs (Sanders and Nafziger, 1976). However, most of these prescriptions are based upon logical analysis rather than upon empirical research. Some of the questions which empirical research might examine are the impact on the information yielded from an evaluation study of various violations, both of degree and of kind, of these prescriptions, the efficacy of alternative procedures for developing and choosing a design, the acceptability of various design types to program personnel, participants and audiences in typical program settings, and the characteristics of various design elements. A great deal of potentially valuable information relevant to design decisions is available in a variety of empirical studies. Smith and Hurst (1974), for example, found no difference in student or parent assessments of a college orientation program when data was collected immediately or after a ten day interval. A study by Carrier, Howard and Miller (1974) carries implications for plans to use student ratings of college courses and instructors. They found that ratings obtained from students attending the last regular class meeting were more favourable than those given by students attending the final exam only and that the sets of ratings did not differ for students who were present on both occasions. A final example refers to the efficacy of the "snowball" method of sampling for obtaining control groups. Palomba (1975) found that a sample obtained in this way in an evaluation study of West Virginia University's mining extension program was well matched to volunteer program participants on several key demographic variables but not on attitudinal variables, providing a warning for evaluation consultants planning to sample in this way.

Evaluation consultants making evaluation process decisions in the area of instrumentation can also turn to an extensive literature for assistance. Key research questions in this area are concerned with the characteristics of a variety of data collection methods for typical populations, including changes in data quality and cost which are associated with changes in such things as question formats, test length, mode of administration, and sample size. In addition, procedures for choosing a data collection method, for choosing an instrument from among those already developed, and for choosing between these two approaches need to be developed and investigated. Publications which provide classified descriptions and reviews of existing data gathering instruments (Borich & Madden, 1977; Buros, 1972; Shaw and Wright, 1967) are a valuable resource in this latter area. There are also many empirical studies which throw light on a variety of data collection methods and procedures for instrument development (e.g. Follman, et al., 1974; Heistad, et al., 1975; Hogben, 1973; Moy, 1972; Nowakowska, 1973). Studies by Kuder (1939) and Straton (1975) have compared the characteristics of alternative methods for obtaining a rank ordering of a set of stimuli while Simpson and Koenig (1975) have investigated the semantic differential for measuring attitudes in the context of drug education research. After examining various ways of assessing teacher effectiveness Glass (1977) rejected the use of student performance measures due to the low or non-significant stability coefficients reported in a number of studies. There is a need for the findings of these and other studies to be made readily accessible to evaluation consultants, particularly studies completed in an Australian context which are particularly difficult to track down.

The final step in the information obtaining process is concerned with data analysis. Here an evaluation consultant is faced with numerous possibilities including data analytic procedures which are univariate or multivariate and parametric or non-parametric. To a large extent the consultant's choice will be determined by the nature of the data at hand and the assumptions underlying the alternative procedures. The appropriateness of these procedures in various situations is well dealt with in textbooks and other sources (Cohen, 1975; Kerlinger and Pedhazur, 1973; Siegel, 1956; Winer, 1971). However, there are still a variety of issues for research in the area of data analysis in evaluation studies. These issues are concerned with the robustness of various data analytic procedures, the development and investigation of new procedures specifically oriented to the problems of data analysis in evaluation, the development and investigation of procedures for choosing a data analysis technique, the costs of alternative analysis procedures and the meaningfulness of various statistics to typical evaluation study audiences. From the results of a study of the relative merits of various measures of average student change for assessing the "true" impact of an educational program Richards (1975) recommended the use of the simplest and most easily understood measure, raw gain scores, in data analysis. A second example of research in this area is a new data analysis procedure with special appeal for use in evaluation studies which has been suggested by Bruno (1972) and tried out by Hardy (1975) who has drawn attention to some limitations to its applicability.

## Research of the Providing Process

The information providing process consists of the documentation and the dissemination of the information gathered during an evaluation study. Research in the area of information documentation might focus upon the characteristics of various media and methods of documentation, the development and investigation of procedures for the optimal matching of audiences and documentation methods, the level of difficulty and sophistication of the documented information which typical audiences can understand, the cost per unit of information and per audience member of various documentation methods, and the time required for production using different methods. Alkin, et al. (1974) have investigated various attributes and correlates of written evaluation study reports of Title VII projects in the area of bilingual education, providing some useful information on documentation. The study by Groundwater-Smith (1975), mentioned above, also has relevance here. She found that audience ratings of the final evaluation study report on eight documentation criteria were similar across five audiences, parents, teachers, principals, school inspectors, and teacher educators. However, the ratings were dissimilar on two other criteria, whether the report was easily read and whether it was possible to form a judgment about the program from the report.

Dissemination research, in the context of evaluation studies, is concerned with investigating alternative ways of delivering the documented information to study audiences. Here dissemination modes might be described and compared in terms of their audience coverage, credibility and acceptability to audiences, cost, and potential for gaining audience attention, as mentioned earlier. However, other questions of concern regarding each mode might be the time and expertise required for dissemination, the flexibility and repeatability of dissemination activities, requirements for special equipment, and the impact on utilization of disseminating information in different orders both within and between audience groups. Empirical research studies in this area are rare. Weiss (1972b, p. 123) refers to studies concerned with dissemination via personal contact which was apparently an effective means of making the results of studies known. The study by Alkin, et al. (1974), mentioned previously, found that Title VII evaluation study reports typically were disseminated to very few audiences. While reports were distributed to the project director and the Federal agency sponsor of virtually every project, the reports were distributed to other audiences in far fewer cases, e.g. teachers (69%), parents (36%), students (2%), school board (79%) and news media (17%).

## Research of Information Utilization

For some time there has been concern that there is a low rate of utilization of evaluation study information (Weiss, 1972b). This concern has stimulated some empirical research of the factors related to information utilization. However, research more specifically oriented to the processes of decision-making and judging is also relevant here. Information utilization research has as its concern how utilization might be conceptualized, the factors affecting utilization, the development and investigation of procedures designed to maximize the likelihood of utilization, and the costs and benefits of utilization and non-utilization in a variety of typical settings. Rolf (1971), in a descriptive study, used several different methods to monitor the variety of dissemination activities following an evaluation study of modular scheduling in a senior high school and record the impact of the study. He revealed several potential types and places of impact. Several other studies have been concerned to discover the factors affecting utilization. Lange and Speiss (1974) found that the legitimacy or credibility of the evaluation consultant was of greatest importance with the frequency of his or her interaction with the decision-maker and the interpretability of evaluation reports also being important. In another study, Goldy (1974) found interpersonal relationships to be significant in discriminating between evaluation studies rated as having high or low value for decision-making. Patton, et al. (1975) also identified what they called the "personal factor", i.e. the interest, enthusiasm and commitment of key audience members, as being the single most important factor affecting utilization. They felt that the evaluation consultant can be influential here by being responsive to the information needs of these key people. Their investigations also suggested that common conceptions of information utilization are too narrow and time bound.

Much of the research of the decision-making and judging processes conducted in fields such as administration, education, health and psychology has implications for the use of evaluation study information in decision-making and judging. It also has implications for the conduct of evaluation studies. The research questions of most concern in evaluation settings are what factors affect decisions and judgments in these settings, what influence do they have, and what procedures may be used to facilitate the making of optimal decisions and judgments. Also of concern are the tradeoffs between costs and other resource usage, on the one hand, and the benefits of using these procedures in typical settings and with typical audiences, on the other. Becker and McClintock (1967) provide a useful review of prescriptive and descriptive decision models and related empirical research while Edwards, Guttentag and Snapper (1975) describe and illustrate a decision-theoretic approach to evaluation studies. Studies by Mehrens (1965) and by Holder and Ehling (1967), while not specifically concerned with evaluation study information, also provide useful insights.

## Research of Evaluation Frameworks and Guidelines

Evaluation study frameworks and guidelines are essentially prescriptive in nature. The frameworks, or models, which have been proposed (e.g. Alkin, 1969; Kourilsky, 1973; Parlett and Hamilton, 1976; Stake, 1967) tend to be broad in scope, providing an overall structure for evaluation studies and suggesting the types of information which might be obtained and the activities and responsibilities of evaluation consultants. Each framework provides a particular perspective without giving detailed prescriptions for the conduct of evaluation studies. Guidelines are usually much more specific, providing detailed step by step guidance for the conduct of an evaluation study (e.g. Center for the Study of Evaluation, 1974; Grotelueschen and Kemmis, 1973; Groundwater-Smith and

Nicholl, 1976; Reineke, 1977). They typically are developed for the evaluation of a particular class of program and may include checklists, data gathering instruments, computational worksheets for data analysis and an outline for an evaluation study final report.

The general nature of evaluation study frameworks makes it difficult for empirical research of their efficacy and utility to yield generalizable information since each framework might be operationalized in a wide variety of ways in different studies. Nevertheless, useful information should come from research which focusses upon describing and comparing alternative frameworks in terms of their applicability and usability in a variety of settings, their adaptability, their credibility and acceptability to various audiences and to program personnel, the types and characteristics of the information which they typically yield, the influence of varying levels of resource allocation and evaluation consultant expertise upon outcomes, and the impact of obtained information upon audience decision-making and utilization. In one of the few comparative studies in this area McAfee (1975) found that the use of a user-oriented framework for the conduct of an evaluation study of a social housing program had several advantages over a more traditional evaluation study which did not provide for citizen participation in the evaluation process. A different approach was taken by Hetke (1975) who studied the content validity of a model for the evaluation of educational administrative services by obtaining the opinions of professors of education, school superintendents, school principals, and lay community leaders regarding various aspects of the model.

Most evaluation study guidelines have been developed for use by persons in a particular field who have had limited training and/or experience in the conduct of evaluation studies. Thus, an important focus for research in this area is the characteristics of evaluation studies conducted using the guidelines for the target group, but also for other likely users. These characteristics might include the acceptability and usability of the guidelines, their appropriateness for use with common programs and in common settings in the field of concern, the quality of the prescribed data gathering instruments, the appropriateness and meaningfulness of the data analysis procedures, the effectiveness of the suggested information documentation and dissemination procedures, evaluation study costs, the time required for the conduct of the evaluation study, and the level of utilization of the information yielded from studies using the guidelines. In a comprehensive experimental study Reineke (1977) allocated 80 curriculum implementation projects to one of two guidelines groups or two control groups. Evaluation studies which used the guidelines were found to be slightly more costly and time consuming according to project director self report, but were judged to be clearer, more complete and more useful by funding agency decision makers. Several other studies have assessed the adequacy of guidelines through ratings obtained from users of the guidelines or an outside panel (Keller, 1974; Mori, 1975; Reinhard, 1972) while Donaldson (1974) compared a long and a short form of a program assessment instrument.

## PRIORITIES AND DIFFICULTIES IN EVALUATION PROCESS RESEARCH

Evaluation process research is potentially a wide and diverse field of activity. Its concerns range from the psychometric properties of various data collection methods to ways of facilitating productive interpersonal interaction and communication between evaluation consultants and program audiences, sponsors and personnel. It is also potentially a very fruitful field since evaluation study information can have a profound influence upon the educational and social programs which society will support and upon their effectiveness. Several authors have recognised the need for empirically based studies of the evaluation process (Sanders and Guba, no date; Straton, 1976; Weiss, 1972b; Worthen and

Sanders, 1973) but little systematic work has been done so far. This is understandable, however, since there are many difficulties or impediments to be overcome in evaluation process research. Much of the information derived from empirical work which is available to assist in evaluation process decision-making relates to the areas of study design, instrumentation and data analysis and is drawn from studies undertaken in other fields. Research specifically oriented to evaluation tends to be concerned with the development and trial of specific frameworks or guidelines and to be of recent origin.

## Priority Areas for Evaluation Process Research

The areas of highest priority for evaluation process research would seem to be those involving an interface between an evaluation consultant and persons concerned with a program including study audiences, study sponsors, program participants and program personnel. These are the areas of the information delineating and the information providing processes. The importance of the "personal factor" was a feature of the findings of several utilization studies (Goldy, 1974; Lange and Speiss, 1974: Patton et al., 1975) and this factor, and its impact on the utility of evaluation studies, needs further clarification. Research might focus specifically upon ways of identifying and prioritizing evaluation study audiences, ways of using audiences as a specific referrent in determining the information to be gathered, ways of tailoring information documentation and dissemination procedures to audience needs, and the impact of alternative approaches in these areas upon information utilization. In addition, further research into the utilization of evaluation study information and the factors affecting utilization should be given high priority.

One further area which should be singled out as being of high priority for evaluation process research is the development and trial of evaluation study guidelines. It will be clear that most evaluations will be undertaken by persons with limited training or experience in the conduct of evaluation studies. Thus, it is important that good quality materials be developed to aid them in their evaluation and decision-making tasks. The guidelines produced by Groundwater-Smith and Nicholl (1976) are a worthwhile beginning in this area. However, there is a need for more detailed guidelines, specifically oriented towards the conduct of evaluation studies of particular types of programs in particular fields, which have been the subject of rigorous trial, as was the case with the guidelines developed by Reineke (1977).

## Difficulties of Evaluation Process Research

The many difficulties associated with any empirical research in the social and behavioural sciences are also difficulties to be faced in evaluation process research. However, there are several further areas of difficulty related to the nature of evaluation studies. A major problem arises from the fact that social and educational programs, in their particular settings, typically differ from each other in such a variety of ways that program replicability or equivalence can rarely be achieved. The circumstances surrounding Reineke's (1977) study were most unusual in that he was able to randomly allocate 80 similar projects to various conditions. All these projects were of very short duration and funded by a single agency in contrast to the usual situation where there is only a single program to be evaluated. A second problem is related to the fact that evaluation process research tends to be atheoretical in nature, at least at present. There is a need for the development and testing of appropriate theory related to the conduct of evaluation studies. A third problem arises from the typically long time interval between the initial decision to undertake an evaluation study and the utilization of the information provided by the study. Thus, the ultimate criterion for comparing alternative approaches to evaluation, that the audiences have made "better" decisions and judgments than they otherwise might have done, will be both remote and non-specific in most studies.

There are many other difficulties of evaluation process research which are of a more practical nature. One of these is that evaluation studies are often initiated only shortly before the program of interest gets underway. Thus, there is little opportunity to plan associated evaluation process research. Second, the concerns of research may be in conflict with the concerns of the evaluation study and may require even more of the time and effort of program participants than would be the case for an evaluation study alone. Third, resources are typically limited for evaluation studies and so additional funds would be required, probably from an additional source, to piggyback evaluation process research on an evaluation study. Other difficulties might be related to the uncertainty of academic recognition for research in this area, the problems of finding time to undertake both an evaluation study and a research study concurrently, and the threats to the research effort posed by political, ethical and other factors which might threaten the evaluation study itself. Evaluation process research requires careful conception and planning if these and other difficulties are to be satisfactorily met.

## CONCLUSION

The quality and utility of any evaluation study will largely be dependent upon the quality of the evaluation process decisions made during the conduct of the evaluation study. The aim of evaluation process research is to develop a basic understanding of the evaluation process and to provide good quality information about each stage of this process and the alternative courses of action which are available. In this way evaluation consultants should be assisted in making more optimal evaluation process decisions. In this paper a model of the evaluation process has been described and used to identify areas for research and to classify and review sample empirical studies which shed light on the evaluation process. Priority areas for research and difficulties of research have also been identified. In one sense, some evaluation process research studies are similar to summative evaluation studies in that they will be oriented towards the decisions faced by a particular audience, evaluation consultants. However, evaluation process research should be broader than this, seeking findings that are widely generalizable and that clarify the evaluation process. This type of research should now be given high priority by persons concerned for the conduct of good quality evaluation studies.

## REFERENCES

Airasian, P.W. Designing summative evaluation studies at the local level. In W.J. Popham (Ed.), *Evaluation in education: Current applications*. Berkeley, California: McCutchan, 1974.

Alkin, M.C. Evaluation theory development. *Evaluation Comment*, 1969, *2* (1), 2-7.

Alkin, M.C., Kosecoff, J., Fitz-Gibbon, C. and Seligman, R. Evaluation and decision making: The Title VII experience. *CSE Monograph Series in Evaluation*, 1974, No.4.

Becker, G.M. and McClintock, C.G. Value: Behavioral decision theory. *Annual Review of Psychology*, 1967, *18*, 239-286.

Borich, G.D. and Madden, S.K. *Evaluating classroom instruction: A sourcebook of instruments*. Reading, Massachusetts: Addison-Wesley, 1977.

Bruno, J.E. A methodology for the evaluation of instruction or performance contracts which incorporates school district utilities and goals. *American Educational Research Journal*, 1972, *9*, 175-195.

Buros, O.K. Mental Measurements Year Books. Highland Park, N.J.: Gryphon Press, 1938, 1940, 1949, 1953, 1959, 1965, 1972.

Campbell, D.T. Reforms as experiments. In E.L. Struening and M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills, California: Sage, 1975.

----, Assessing the impact of planned social change (Paper #8 in Occasional Paper Series). Kalamazoo, Michigan: Western Michigan University, Evaluation Center, December, 1976.

Carrier, N., Howard, G.S. and Miller, W.G. Course evaluation: When? Journal of Educational Psychology, 1974, 66, 609-613.

Center for the Study of Evaluation, Summative evaluation kit. Los Angeles: Center for the Study of Evaluation, U.C.L.A., 1974.

Cohen, J. Multiple regression as a general data-analytic system. In E.L. Struening and M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills, California: Sage, 1975.

Donaldson, B.L. Development and evaluation of an instrument for the qualitative evaluation of classroom programs for the emotionally impaired. Unpublished doctoral dissertation, Michigan State University, 1974.

Edwards, W., Guttentag, M. and Snapper, K. A decision-theoretic approach to evaluation research. In E.L. Struening and M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills, California: Sage, 1975.

Follman, J., Wiley, R., Geiger, G. and Lavely, C. Effects of halo-effect instructions on level of ratings and intercorrelations. Psychological Reports, 1974, 35, 392.

Glass, G.V (Ed.), Evaluation studies review annual (Vol. 1). Beverly Hills, California: Sage, 1976.

Glass, G.V A review of three methods of determining teacher effectiveness. In G.D. Borich (Ed.), The appraisal of teaching: Concepts and process. Reading, Massachusetts: Addison-Wesley, 1977.

Goldy, F.B. Program evaluation as a problem of organizational interaction. Unpublished doctoral dissertation, University of Southern California, 1974.

Grotelueschen, A. and Kemmis, S. The evaluation of library demonstration projects. Unpublished manuscript, Center for the Study of Information and Education, Syracuse University, June, 1973.

Groundwater-Smith, S. A formative evaluation of a vertically grouped two-teacher open space unit. Unpublished B.A. (Honours) thesis, University of Sydney, 1975.

Groundwater-Smith, S. and Nicoll, V. Tell it how it is: Some guidelines for project evaluation. Sydney: Standard Publishing House, 1976.

Guttentag, M. and Struening, E.L. (Eds.), Handbook of evaluation research (Vol. 2). Beverly Hills, California: Sage, 1975.

Hardy, R. A limitation of the applicability of interval shift analysis to program evaluation. American Educational Research Journal, 1975, 12, 544-549.

Heisted, G.T., Zimmermann, R.L. and Wong, M.R. Measuring changes in attitude toward drug abuse: A preliminary report of a method. Journal of Drug Education, 1975, 5 (2), 127-140.

Hetke, R.L. A study of a model for the evaluation of educational administrative services. Unpublished doctoral dissertation, Northern Illinois University, 1975.

Hogben, D. The reliability, discrimination and difficulty of word-knowledge tests employing multiple choice items containing three, four, or five alternatives. Australian Journal of Education, 1973, 17, 63-68.

Holder, H.D. and Ehling, W.P. Construction and simulation of an information-decision model. Journal of Communication, 1967, 17, 302-315.

Keller, L.J. An appraisal of the proposed evaluative criteria optional instrument. Unpublished doctoral dissertation, Indiana University, 1974.

Kelling, G.L., Pate, T., Dieckman, D. and Brown, C.E. The Kansas City preventive patrol experiment: A summary report. In G.V Glass (Ed.), Evaluation studies review annual (Vol. 1). Beverly Hills, California: Sage, 1976.

Kerlinger, F.N. and Pedhazur, E.J. *Multiple regression in behavioral research*. New York: Holt, Rinehart and Winston, 1973.

Kourilsky, M. An adversary model for educational evaluation. *Evaluation Comment*, 1973, *4* (2), 3-6.

Kuder, G.F. The stability of preference items. *Journal of Social Psychology*, 1939, *10*, 41-50.

Lange, R.R. and Speiss, M. *An empirical study of four providing variables effecting utilization of evaluation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1974.

McAfee, R.A. *Interactive evaluation: A user-oriented process to assist housing program reformulation*. Unpublished doctoral dissertation, University of British Columbia, 1975.

Mehrens, W.A. *The use of probability information in educators' decision making*. Unpublished doctoral dissertation, University of Minnesota, 1965.

Mori, A.A. *The development and utilization of an evaluation device for instructional materials completed by subscribers to the Western Pennsylvania Special Education Regional Resource Center*. Unpublished doctoral dissertation, University of Pittsburgh, 1975.

Moursund, J.P. *Evaluation: An introduction to research design*. Belmont, California: Wadsworth, 1973.

Moy, M.L.Y. *Item sampling: Optimum number of people and items*. Unpublished doctoral dissertation, Ohio University, 1972.

Nevo, David *Evaluation priorities of students, teachers and principals*. Unpublished doctoral dissertation, Ohio State University, 1974.

Nowakowska, M. Perception of questions and variability of answers. *Behavioral Science*, 1973, *18*, 99-108.

Nunnally, J.C. The study of change in evaluation research: Principles concerning measurement, experimental design and analysis. In E.L. Struening and M. Guttentag (Eds.), *Handbook of evaluation research* (Vol. 1). Beverly Hills, California: Sage, 1975.

Palomba, C. "Snowball" and other control groups. *Journal of Vocational Behavior*, 1975, *6*, 175-184.

Parlett, M. and Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programmes. In D.H. Tawney (Ed.), *Curriculum evaluation today: Trends and implications*. London: Schools Council Research Studies, MacMillan, 1976.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., French, B.D. and Blyth, D.A. *In search of impact: An analysis of the utilization of federal health evaluation research*. Minneapolis: Minnesota Center for Sociological Research, 1975.

Popham, W.J. *Educational evaluation*. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Reineke, R.A. *The development and pilot test of an evaluation guide for directors of a class of educational projects*. Unpublished doctoral dissertation, University of Minnesota, 1977.

Reinhard, Diane Lois *Methodology development for input evaluation using advocate and design teams*. Unpublished doctoral dissertation, Ohio State University, 1972.

Richards, J.M. A simulation study of the use of change measures to compare educational programs. *American Educational Research Journal*, 1975, *12*, 299-311.

Rippey, R.M. (Ed.), *Studies in transactional evaluation*. Berkeley, California: McCutchan, 1973.

Rolf, S. *A follow-up of the White Bear Lake Senior High School evaluation by the E.R.D.C.* Unpublished manuscript, University of Minnesota, Department of Psychological Foundations of Education, 1971.

Sanders, J.R. and Guba, E.G. *An examination of factors affecting the practice of educational evaluation*. Unpublished manuscript, Western Michigan University, Evaluation Center, No date.

Sanders, J.R. and Nafziger, D.H. *A basis for determining the adequacy of evaluation designs* (Paper #6 in Occasional Paper Series). Kalamazoo, Michigan: Western Michigan University, Evaluation Center, April, 1976.

Scriven, M.C. Evaluation perspectives and procedures. In W.J. Popham (Ed.), *Evaluation in education: Current applications*. Berkeley, California: McCutchan, 1974.

Shaw, M.E. and Wright, J.M. *Scales for the measurement of attitudes*. New York: McGraw-Hill, 1967.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Simpson, M.L. and Koenig, F.W. The use of the semantic differential technique in drug education research: An example and some suggestions. *Journal of Drug Education*, 1975, *5* (3), 251-259.

Smith, T.T. and Hurst, J.C. Immediate follow-up in orientation program evaluation. *Journal of College Student Personnel*, 1974, *15*, 46-48.

Smith, W.G. Evaluation of the clinical services of a regional mental health center. In G.V Glass (Ed.), *Evaluation studies review annual* (Vol. 1), Beverly Hills, California: Sage, 1976.

Spivack, G., St. Clair, C.H., Siegel, J. and Platt, J.J. Differing perspectives on mental health evaluation. *American Journal of Psychiatry*, 1975, *132*, 1295-1299.

Stake, R.E. The countenance of educational evaluation. *Teachers College Record*, 1967, *68*, 523-540.

Straton, R.G. Obtaining paired comparisons data from multiple rank orders using partially balanced incomplete block designs. *Educational and Psychological Measurement*, 1975, *35*, 555-566.

----, *Ethical issues in evaluating educational programs*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Brisbane, November, 1976.

----, *Alternative models for school program evaluation*. Paper presented at a School Evaluation Forum of the A.C.T. Teachers Federation and the A.C.T. Schools Authority, Canberra, March, 1977.

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O. and Provus, M.M. *Educational evaluation and decision making*. Itasca, Illinois: F.E. Peacock, 1971.

Taylor, D.B. Eeny, meeny, miney, meaux: Alternative evaluation models. *North Central Association Quarterly*, 1976, *50*, 353-358.

Ten Brink, T.D. *Evaluation: A practical guide for teachers*. New York: McGraw-Hill, 1974.

Weiss, C.H. *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, N.J.: Prentice-Hall, 1972a.

----, Utilization of evaluation: Toward comparative study. In C.H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education*. Boston: Allyn and Bacon, 1972b.

Winer, B.J. *Statistical principles in experimental design*. (2nd edition.) New York: McGraw-Hill, 1971.

Worthen, B.R. and Sanders, J.R. *Educational evaluation: Theory and practice*. Worthington, Ohio: Charles A. Jones, 1973.

## NOTE

[1] The author is indebted to L. Bolitho for assistance with library research related to this paper.

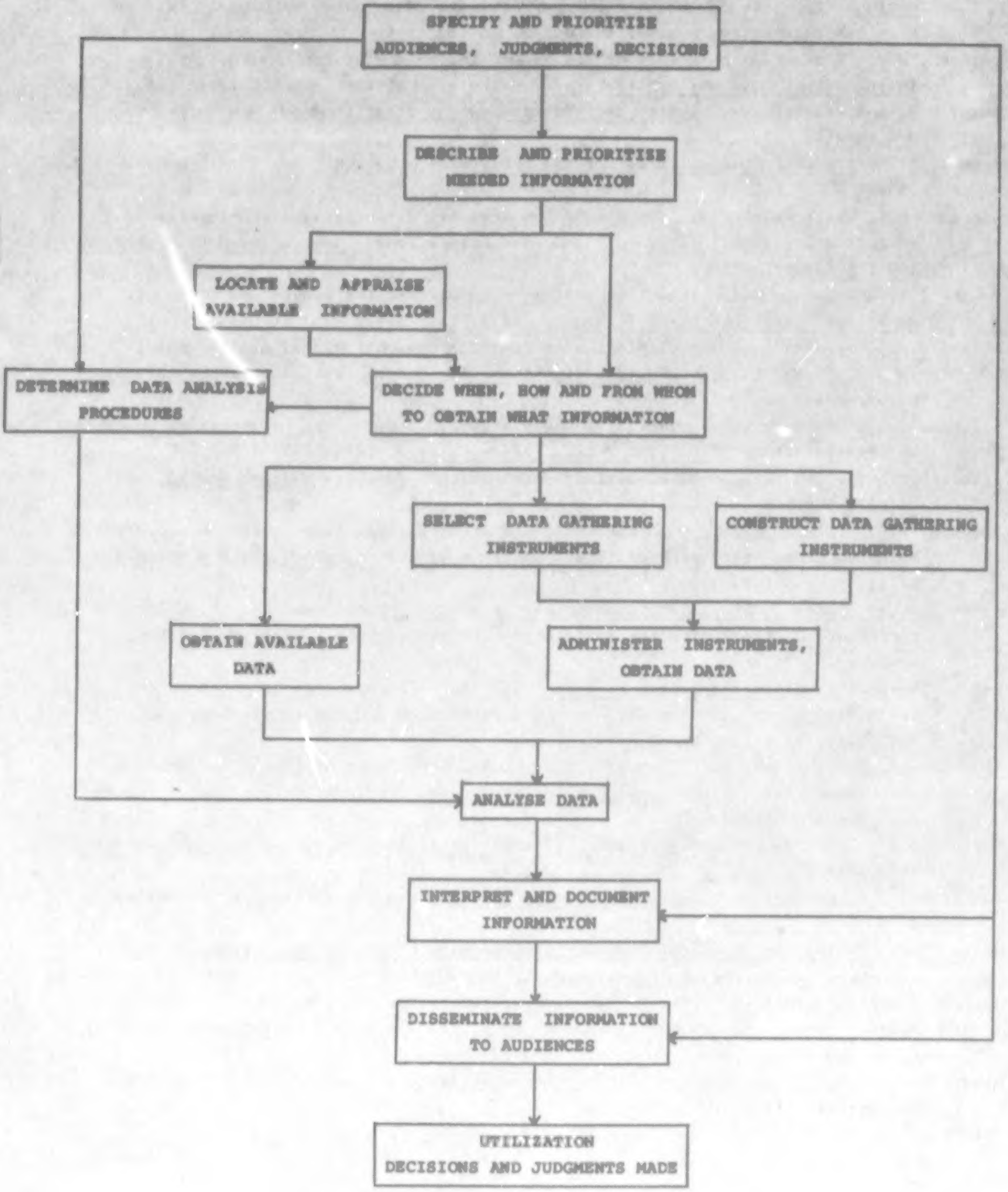

SPECIFY AND PRIORITIZE AUDIENCES, JUDGMENTS, DECISIONS
DESCRIBE AND PRIORITIZE NEEDED INFORMATION
LOCATE AND APPRAISE AVAILABLE INFORMATION
DETERMINE DATA ANALYSIS PROCEDURES
DECIDE WHEN, HOW AND FROM WHOM TO OBTAIN WHAT INFORMATION
SELECT DATA GATHERING INSTRUMENTS
CONSTRUCT DATA GATHERING INSTRUMENTS
OBTAIN AVAILABLE DATA
ADMINISTER INSTRUMENTS, OBTAIN DATA
ANALYSE DATA
INTERPRET AND DOCUMENT INFORMATION
DISSEMINATE INFORMATION TO AUDIENCES
UTILIZATION DECISIONS AND JUDGMENTS MADE

Figure 1. A model of the evaluation process

## I. THE DELINEATING PROCESS

1. AUDIENCES: Audience identification and prioritization methods; Impact of audience identification on information utilization.

2. PURPOSES: Identification and prioritization methods for decisions and judgments; Impact of decision and judgment identification on information utilization.

3. INFORMATION: Information identification, prioritization and specification methods; Valued information for typical audiences and programs.

## II. THE OBTAINING PROCESS

4. INVENTORY: Identification and appraisal methods for existing records; Characteristics of typical records.

5. DESIGN: Procedures for design development and choice; Acceptability of design types to audiences and participants; Impact on utilization of design types; Costs of typical designs.

6. INSTRUMENTATION: Characteristics of data collection methods for typical populations; Procedures for choosing a method and for choosing between available instruments; Costs of methods.

7. ANALYSIS: Characteristics and robustness of techniques, statistics; Procedures for choosing techniques; Costs of techniques; Meaningfulness of alternative statistics to audiences.

## III. THE PROVIDING PROCESS

8. DOCUMENTATION: Characteristics of media and documentation modes; Audience-documentation matching procedures; Costs and production time for documentation modes; Information utilization and documentation modes.

9. DISSEMINATION: Characteristics of dissemination modes; Audience-dissemination mode matching; Costs and time for dissemination modes; Information utilization and dissemination modes and order of dissemination to audiences.

## IV. INFORMATION UTILIZATION

10. UTILIZATION: Conceptualization; Factors affecting utilization; Procedures for increasing utilization.

11. DECISION-MAKING/JUDGING: Descriptive and prescriptive models; Factors affecting decision-making, judging; Procedures for optimizing decision-making, judging.

V. EVALUATION FRAMEWORKS AND GUIDELINES

12. FRAMEWORKS: Applicability, usability, acceptability; Characteristics and types of information and impact on information utilization.

13. GUIDELINES: Applicability, usability, acceptability appropriateness in common settings; Appropriateness and effectiveness of data gathering, analysis, documentation, dissemination procedures; Time and cost requirements; Information utilization levels.

Figure 2. A classification of evaluation process research

## FRIDAY 11 NOVEMBER

## 3.00 - 4.00pm

| PAPERS | Rooms |
|---|---|
| D. Hogben (Flinders University<br>Curriculum Evaluation: By Whom, For Whom? | Seminar Room A |
| OMITTED<br>J. Cotterell (University of Queensland)<br>Environment Contexts and Student Satisfactions<br>(Paper will be distributed at Conference) | Ladies Drawing Room |
| P. Sheldrake (Flinders University)<br>Curriculum Evaluation and Objectivity: Problems in Policy-Oriented Research | Seminar Room D |
| F.J. Hunt (Monash University)<br>Curriculum Evaluation and the Development of the Person | Meeting Room |
| S. Bourke, G. Morgan (A.C.E.R.)<br>The Use of Criterion - Referenced Testing in Curriculum Evaluation | Gateway Room |
| OMITTED<br>TAFE Workshop (2.00-4.00pm)<br>Chairman, R. Catts, N.S.W. Department of TAFE)<br>J. Malley (R.M.I.T.)<br>A Case Study of Needs Evaluation<br>R. Schilling (S.A. Dept. F.E.)<br>A Case Study of Needs Evaluation<br>D. Kuhl (S.A. Dept. F.E.)<br>Course Evaluation - Rationale<br>P. Edwards (N.S.W. P.T.C.)<br>Discussion Opener | Seminar Room B |

# CURRICULUM EVALUATION: BY WHOM, FOR WHOM?

Donald Hogben
Flinders University

Ten years ago Michael Scriven urged educators to distinguish clearly between the *goals* of educational evaluation and the many various *roles* that evaluation can take on. He pointed out that, while the principal *goal* of any evaluation activity is to answer questions about merit or worth, "... the *role* which evaluation has in a particular educational context may be enormously various ..." (Scriven, 1967, p.40).

In the course of his discussion, Scriven pointed to the fact that not only was there confusion between roles and goals, but also that this confusion was tolerated, perhaps even encouraged, in an attempt to lessen anxiety over evaluation. That confusion and anxiety still exist today hardly needs arguing, and I believe that many of the problems and much of the prevailing confusion are due to a lack of clarity regarding the roles that evaluators are expected to play and do play. While we may well agree with Scriven that the ultimate goal of any evaluation activity is the establishment of worth, we are still left with such questions as: why is it important to establish worth in this particular setting at this particular time; for whom is the evaluation information being obtained; who is to carry out the evaluation; how is the evaluation information to be used?

In his 1967 paper, Scriven distinguished two major evaluation roles, the *formative* and the *summative* and these notions are now well embedded in evaluation discourse. In a revised version of the paper, Scriven (1973) additionally identified a hybrid evaluation case which he called *monitoring*. In the *formative* frame the evaluation information (Scriven called it the "evaluation feedback loop" in his revised paper) stays *within* the program, school, or whatever, with the idea of improving it. The evaluation information services decision-making within the program and there is no necessary reporting outside: Stufflebeam (1975) has described this sort of evaluation as *proactive*. With *summative* evaluation, the information goes *outside* to program sponsors, other educators, parents, etc. for some action. Stufflebeam (1975) points out that evaluation for accountability calls for a retrospective approach, and he observes that his "*retroactive* evaluation to serve accountability" is close to Scriven's concept of *summative* evaluation. In Scriven's hybrid case of *monitoring* "... the loop goes outside from an intermediate stage, and the decision serviced usually concerns intervention and support." (Scriven, 1973, p.63)

Whether it is more useful to think of evaluation as *formative*, *summative*, *proactive*, or *retroactive*, or whether other terms might be more valuable, need not concern us further here. However, from classifications such as those of Scriven and Stufflebeam we can identify two categories of persons interested in and/or involved in any given evaluation exercise. For our present purposes it may be profitable to look at our two questions - evaluation by whom, for whom? - through a focus on these two groups which I shall call *insiders* and *outsiders*. By *insiders* I mean the people involved in and responsible for the program, project, curriculum, or whatever it is that is the subject of a particular evaluation. *Outsiders* are those external to the program or project, but who have a legitimate interest in the evaluation of it. *Outsiders* would include funding agencies, school councils, parents, government departments, and external evaluators.

Combining our two questions - evaluation by whom? and evaluation for whom? - and employing the notion of *insiders* and *outsiders*, we can identify *four basic evaluation situations*. These are: evaluation by insiders for insiders; evaluation by insiders for outsiders; evaluation by outsiders for insiders; and, evaluation by outsiders for outsiders. Combinations would not be uncommon of course; for example, evaluation to serve both insiders and outsiders.

I believe it is important for evaluators and project or curriculum personnel to recognize clearly the situation obtaining in each particular case, and the issues and problems that may be inherently present, or that may likely arise in the course of the evaluation. Consumers of evaluation reports too need to be sensitive to the sorts of problems and difficulties likely to accompany each particular situation so that they can better interpret and judge the validity and usefulness of the information and conclusions contained in reports. Educational researchers with an interest in the evaluation process should also be interested in the sorts of problems and issues that I shall touch on in the remainder of this paper, since very few of the conclusions and generalizations that I consider have actually been the subject of careful empirical study.

Since it is obviously not possible to consider, for each of the four situation, the whole range of issues and problems that exist, I have selected the few which seem to me to be particularly important and interesting, and which may serve to stimulate discussion and debate.

## Evaluation by Insiders for Insiders

There are many evaluation undertakings that would clearly fall into this category or situation. Perhaps the simplest, and the one that comes most readily to mind, would be the sort of informal evaluation conducted by teachers on their own courses. Here the person conducting the evaluation and the audience for the evaluation are one and the same; that is, the evaluation is simple *self-evaluation*. It may involve little more than the teacher's own subjective appraisal of his or her efforts and those of the students. Self evaluations can be more elaborate of course and may involve the use of carefully designed testing instruments, student opionionnaires, observation records, etc. On a much larger scale, but still within the same category, we can think of a particular school faculty (or a whole school for that matter) deciding to conduct evaluations of its programs or curricula.

In both the above examples the impetus for the evaluation comes from the *insiders* themselves, and there is no obligation to report the evaluation findings to *outsiders*. An obvious variant would be where the evaluation is not self-initiated but is undertaken at the request, or on the order, of some person or group outside. For example, a principal or a department head may require that all teachers or instructors under his charge regularly conduct course evaluations but that they need not furnish him with any evaluation reports.

Two likely problems with unsupplemented self-evaluations are lack of evaluation expertize and resources (particularly time), and the unreliability of the results. So long as evaluation is self-initiated and motivated by a genuine desire to improve course quality, there seems little reason to suppose that the evaluation will be deliberately falsified in any way. Where self-evaluation is *required*, there may be a distinct lack of enthusiasm for it but deliberate bias or distortion again hardly seems likely if the results of such evaluations need not be communicated to outsiders. However, there seems to exist a common unease about self-evaluation (see, for example, Glass, 1975). Scriven (1975) appears to be in no doubt about the lack of reliability inherent in self-evaluation when he says (p.22): "Since it is a universal truth that self-evaluation is unlikely to be reliable, it is a necessary consequence of interest in truth that one supplement self-evaluation. Hence anyone interested in improving his or her own performance *must* arrange for or endorse some kind of independent evaluation."

Even if we are not prepared to accept fully Scriven's strong assertion, it seems not unreasonable nevertheless to suggest that any individual or group, planning to implement evaluation that is essentially self-evaluation, should obtain supplementary input from some independent outside source whenever possible. Since the independent evaluator will be communicating his findings only to the insiders themselves, there seems little reason to suppose that honest comments and judgments should not be welcomed and taken note of. It doesn't necessarily follow, of course, that they will be enjoyed, as Scriven notes. I shall touch on this point again later in the paper.

### Evaluation by Insiders for Outsiders

This evaluation situation shares much in common with the one I have just looked at; however, it differs in several quite important ways. As in the first situation, the evaluation is essentially self-evaluation, but in this case the evaluation information goes *outside*. In Scriven's terminology it is *summative* evaluation or, perhaps, *monitoring*. Since it is self-evaluation, the problems inherent in self-evaluation that I have just touched on apply and need not be repeated. Often the setting is, however, much more complex and there is a host of additional issues and difficulties that bear looking at.

The situation of evaluation by insiders for outsiders occurs quite commonly with government-sponsored education programs of the sort where large numbers of fairly small projects controlled by individuals or small groups receive short-term financial backing. Two such programs with which I have some familiarity are the Australian Schools Commission's Innovations Program, begun in 1974, and Title III of the Elementary and Secondary Education Act (ESEA) which came into force in the United States in 1965. It is not my intention to consider either of these programs as such, nor is it my intention at any point in this paper to look into the question of *why* particular evaluation situations exist. Nevertheless, I shall concentrate to some extent on this type of project-program arrangement in this section as it brings out so many of the important issues arising from the evaluation by *insiders* for *outsiders*.

It is invariably a requirement of such programs that individual project directors submit evaluation reports to the funding agencies, and sometimes a certain proportion of grant money is earmarked for the financing of project evaluations. In situations where insiders are required to report on their projects to outsiders, it is not uncommon to encounter anxiety, confusion, and sometimes hostility. The first, and perhaps the most obvious confusion is over evaluation *role*. Project personnel are advised to gather evaluation data on a regular basis and to employ these data in the development and improvement of their projects; that is, project personnel are cast in the role of *formative* evaluators. Few insiders object to this, and the majority appear readily to appreciate the value of data-gathering and record-keeping for the purpose of detecting problems and correcting them; discovering strengths and pay offs and building on them. However, as they are also required to report on their activities, often both during and at the conclusion of the funding period, they must also play the role of *monitoring* and *summative* evaluators. The precise purpose of the required report writing and the use funding agencies make of the submitted reports is rarely stated in any sort of detail, but accountability and dissemination are the two most obvious purposes.

It is likely that the funding agencies themselves will be rather ambivalent about project evaluation in the summative or accountability sense and somewhat sensitive on the report issue, since, having funded projects in the belief that they will be successful and worthwhile, sponsors naturally look toward favourable outcomes. On this point Scriven (1975, p.3) notes that the funding agency is in a parental role and that project evaluation is partly an evaluation of the funding body itself. Reporting for dissemination *can* be a very valuable activity, though the transferability and generalizability of educational programs and innovations appears to be a complex business (see, for example, House, 1975; Stake, 1969; MacDonald, 1975).

House (1974) has drawn attention to the *ideologies* that grow out of funded projects and which have important implications for evaluation. He notes that the ideology is normally presented in the project proposal and that "... it is absolutely indispensable in securing the project funds because it says what the sponsors want to hear (p.619)." House goes on to suggest that a close look at ideologies will show that they all contain much the same promises and that there is really no way in which the project can live up to its ideology. The high level of competition for limited funds, plus the added status which tends to accompany a successful application, almost guarantees that applicants will promise more than they are likely to be capable of delivering. If this situation generally obtains, project directors must invariably find themselves faced with the spectre of reporting that they have fallen short of goal achievement. The project director who is aware of this situation, and who is not prepared to report to outsiders that his project has been less than totally successful, can of course omit negative findings from his report or he can avoid gathering data that are likely to be other than supportive.

Several School Commission innovations project directors have expressed the view to me that negative evaluations should never be made. One school principal put his opinion strongly in the following words (or words very close to them): "If principals think that they are going to be blamed in some way if they make a bit of a mess of things, then very few of them will bother applying for grant money. People should be rewarded for 'having a go': they should not be criticized if things go wrong." This principal has in effect argued against any serious evaluation and outside reporting being undertaken, since true evaluation must look for and record project minuses as well as plusses.

House (1974) points to the importance, in the United States at least, of innovative projects for people pursuing careers, and he claims that such projects serve as primary vehicles for promotion and advancement. He goes on to say (pp.624-625) "... not only does the upwardly mobile person need a project with which to build his reputation but the project must be without blemish. Within our system as it operates the *appearance* of the project is as important as the project itself ... Making ... blemishes public is to damage the administrative entrepreneur's career."

Are these things true for Australia? If they are, and if many of the funded projects in this country are being used in the manner House suggests, then the validity of the reports based on evaluations by insiders must be suspect. Quite obviously not all project personnel will be preoccupied or even concerned with these things. However, project directors and others interested, particularly the consumers of evaluation reports at all levels, should be aware of the likelihood of distortions: deliberate or unwitting.

I have concentrated somewhat in this section on funded project self-evaluation. However, before moving on to the next section, I shall touch on one other circumstance: I refer to the case where no evaluation is requested or required, but a teacher, a faculty, perhaps a whole school decides to conduct a self-evaluation for the purpose of reporting to some outside person or group. One can imagine evaluations of this kind being undertaken for a variety of reasons, and I venture to suggest that the motivation for conducting such evaluations will bear strongly on the manner of their conduct and the ultimate credibility of any reports emanating from them.

Evaluations sponsored internally are no doubt more likely to occur when program or curriculum personnel believe that things are going along well and when it is seen to be advantageous if this fact is known outside. I have previously suggested (Hogben, 1975), that internally-initiated evaluations may be conducted in order to justify current practices and/or to counter actual or anticipated challenges from outside. Challenges from superiors, from pressure groups, politicians, the media, may be effectively met and turned aside by favourable evaluation reports. Evaluations initiated by insiders for the sole purpose of defending the *status quo* need not necessarily be totally lacking in credibility, but they are hardly likely to involve a diligent search for program shortcomings.

### Evaluation by Outsiders for Insiders

In this situation we have outsiders - presumably people with some evaluation expertize - being called upon to conduct, or assist with, evaluations. The involvement of outside evaluators can range from simple situations of the kind previously touched on, where a teacher conducting an evaluation of his or her own course(s) invites an interested and competent colleague to provide some independent assessments, through to the employment of full-time formative evaluators on large scale curriculum development projects. No matter whether the situation is simple or complex, small or large scale, the same general issues emerge and much the same problems seem to present themselves.

Perhaps the most obvious, and certainly the most often mentioned of these is the danger of the external formative evaluator(s) becoming co-opted.

In recent years there has been a fairly strong movement within the evaluation field by a number of evaluators and evaluation theorists away from what Parlett and Hamilton (1972) have referred to as the "agricultural-botany" paradigm. Hastings was an early writer on this topic and he anticipated some of the moves in his 1969 paper titled "The Kith and Kin of Educational Measurers". In a recent paper, Eisner (1977), like Hastings, advances the view that the evaluation of education has been too parochial in the past and that the "scientific approaches" should be complemented. He argues that the concepts of educational connoisseurship and educational criticism - concepts embedded in an artistic tradition - yield procedures that could prove valuable. Whether Eisner's ideas find a place in the general evaluation field remains to be seen. Meanwhile, the movement away from evaluation strategies and approaches more-or-less firmly embedded within the psychometric tradition, has resulted in several approaches which have much in common. I refer to approaches such as "responsive evaluation" (Stake, 1975); "illuminative evaluation" (Parlett and Hamilton, 1972); "transactional evaluation" (Rippey, 1973);and "democratic evaluation" (MacDonald, 1976). Since all of these evaluation approaches place considerable stress on the interaction between evaluators and program or curriculum personnel, there is the ever-present problem of the evaluator being so influenced as to lose his objectivity and independence.

The problem has not gone unrecognized by the proponents of such approaches, and it is specifically referred to by at least Parlett and Hamilton and Stake. Co-option and loss of independence are not problems that are peculiar only to these approaches, of course. Scriven, for example, has written extensively and forcefully on the general issue of evaluator co-option and the need for evaluators to maintain their independence, (Scriven, 1967, 1972, 1975). Scriven's now well-known "goal-free" evaluation strategy was developed at least in part as an attempt to deal with this sort of issue in both formative and summative situations (Scriven, 1971).

A problem which in a sense is the opposite of that just discussed concerns the threat which evaluation may present. Even the most sympathetic of evaluators and the most scrupulously fair evaluation plans and procedures pose some threat, at least initially, to the person or group whose efforts are being scrutinized. A number of people have written of the need for evaluators to gain the confidence of project personnel and to seek their cooperation in the conduct of formative evaluations (see, for example, Harlen, 1975).

Evaluation-induced anxiety is not likely to be generally serious in situations where outside evaluators communicate their findings only to the insiders who engaged them: the threat potential is far greater when outside reporting is called for. Nevertheless, curriculum or program personnel can be extremely sensitive to critical comment. Carter (1971) has addressed the issue of client resistance to negative evaluation results and in the opening paragraph of his paper he observes (p.80): "This resistance unveils one of the primary latent functions of many evaluation studies, namely, a legitimization process for predetermined policies."

In the course of the paper, Carter puts forward three propositions relating to response to negative findings: "The greater the perceived threat to the client's or manager's positive self-concept, the greater will be his resistance to negative research findings (p.85)." "The greater the difference between the client's or manager's concept of the social reality being studied and the research findings, the greater will be his resistance to the results (p.85)." "The greater the importance or salience to the client or manager of the function being evaluated, the greater the perceived threat to his positive self-image and the greater his resistance to the negative findings (p.87)."

Several years ago House and I conducted a limited goal-free formative evaluation of a special curriculum project for a well-known and highly regarded curriculum group in the United States. We had some rather critical things to say and the reactions to our criticisms were very revealing, and appeared to confirm the sorts of propositions advanced by Carter. Our study and our critical comments were generally well-received by the curriculum development team members - the people actually engaged in the day-to-day business of developing materials and trialling them. The response from "top management" was, however, quite negative and in one case down right hostile. Our use of the new and then quite revolutionary "goal-free" approach (all this took place in 1972) of course made us rather more vulnerable than had we adopted a more "conventional" methodology.

## Evaluation by Outsiders for Outsiders

This evaluation situation is possibly the one with which most people are most familiar. It is certainly the situation about which much has been written, and for that reason I shall not dwell long on it here in what is perhaps an already-too-long paper.

In each of the other three evaluation situations that I have look at, the insiders featured in some way. Here, the individuals or groups whose activities are being examined are neither the primary audience for evaluation findings nor are they necessarily involved in the planning and conduct of the evaluation. In the other situations we have seen that the inside and the outside groups are usually readily identifiable, and the evaluation motivations fairly obvious. With evaluation by outsiders for outsiders, however, the situation can be much more involved; evaluation motivations are often somewhat obscure, and the potential threat to those being evaluated is frequently high. Of course the stakes too may be high; for example, evaluations by accrediting teams ("blue ribbon panels") site visiting schools or other educational institutions. Nevertheless, school people generally may perhaps be excused for a lack of enthusiasm for this form of evaluation which is clearly not primarily intended to be of immediate benefit to them and their students.

It has often been pointed out that evaluation can be a highly political exercise; in fact some writers have suggested that evaluation is always political (Cohen, 1970). It seems particularly important to examine motives, and carefully to look at relationships existing between different individuals and groups, in those evaluation situations in which outsiders are providing themselves with information or gathering it on behalf of other outsiders. This is not to suggest that all such evaluations are born out of sinister motives. Many are quite obviously straight-forward and fairly neutral. However, on many occasions even a rather cursory examination may reveal them to be highly political: the aim being perhaps to justify, perhaps to discredit.

The formal relationships existing between evaluators, program sponsors, evaluation sponsors, and those whose activities are the subject of study, frequently appear to be of considerable significance. For example, if the outside evaluator is in the permanent employ of those sponsoring the program, or if he is obligated to them in some other way, his credibility and that of his findings is unlikely to be as high as would be the case if he were totally independent; particularly if he retains certain rights regarding his findings. Scriven (1975) has written specifically and quite extensively on the issue of evaluation bias and ways in which it may be overcome or controlled. Others (House, for example) in the United States have written a good deal on the general question of the politics of evaluation, and particularly the issue of accountability and its attendant problems.

Recently in Britain, MacDonald (1976) has suggested a political classification of evaluation studies, and he has identified three types which he labels bureaucratic, autocratic, and democratic. Classifying evaluators and evaluation studies in this way may prove to be of considerable help in the identification of issues, and in the anticipation of problems and difficulties likely to arise with evaluation by outsiders. We might note, though, that MacDonald's democratic evaluation would cover evaluation both for insiders and outsiders. It shares a good deal with Stake's responsive evaluation.

## Conclusion

In this paper I have identified four basic evaluation situations based upon the varying roles that might be played by curriculum or program insiders and outsiders. I have explored some of the issues and problems that tend to be associated with each of these different evaluation settings, and I have suggested that closer attention to them may help to overcome, or at least to reduce, much of the ambiguity and uncertainty that currently prevails. In the future, very careful attention to the two important questions "evaluation by whom, for whom?" should better alert us to the issues and problems that will have to be faced either before or during the life of any particular evaluation.

Unfortunately, there is very little good empirical research evidence presently available to guide our decisions and our practice. I hope that we can look forward to an increased research interest in the evaluation process itself in the near future. Maybe some of the points I have raised in the course of this paper will provide some focus.

## REFERENCES

Carter, R.K. Client's Resistance to Negative Findings. American Sociologist, 1971, 6, 118-124. Reprinted in House, E.R. (Ed.), School Evaluation. Berkeley: McCutchan, 1973, 80-95.

Cohen, D.K. Politics and Research. Review of Educational Research, 1970, 40, 213-238. Reprinted in House, E.R. (Ed.) School Evaluation. Berkeley: McCutchan, 1973, 96-124.

Eisner, E.W. On the Uses of Educational Connoisseurship and Criticism for Evaluating Classroom Life. Teachers College Record, 1977, 78, 345-358.

Glass, G.V A Paradox About Excellence of Schools and the People in Them. Educational Researcher, 1975, 4, 9-13.

Harlen, W. Science 5-13: A Formative Evaluation. Schools Council Research Studies: Macmillan, 1975.

Hogben, D. School Evaluation: Motivations and Outcomes. Australian Journal of Social Issues, 1975, 10, 111-119.

House, E.R. The Politics of Evaluation in Higher Education. Journal of Higher Education, 1974, 45, 618-627.

House, E.R. Transferability and Equity in Innovation Policy. Paper presented at the National Conference on Innovation and Change, Detroit, Michigan, September 1975.

Hastings, J.T. The Kith and Kin of Educational Measurers. Journal of Educational Measurement, 1969, 6, 127-130.

MacDonald, B. Humanities Curriculum Project. In Evaluation in Curriculum Development: Twelve Case Studies. Schools Council Research Studies: Macmillan, 1975, 80-90.

MacDonald, B. Evaluation and the Control of Education. In Tawney, D. (Ed.), Curriculum Evaluation Today: Trends and Implications. Schools Council Research Studies: Macmillan, 1976, 125-136.

Parlett, M. & Hamilton, D. Evaluation as Illumination: a New Approach to the Study of Innovatory Programs. CRES, University of Edinburgh, Occasional Paper No. 9, 1972.

Rippey, R.M. (Ed.). Studies in Transactional Evaluation. Berkeley: McCutchan, 1973.

Scriven, M. The Methodology of Evaluation, American Educational Research Association Monograph Series on Curriculum Evaluation, Chicago: Rand McNally, 1967, 1, 39-83.

Scriven, M. The Methodology of Evaluation. In Worthen, B.R. & Sanders, J.R. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.

Scriven, M. Goal-free Evaluation. An unpublished working paper for the National Institute of Education, November, 1971.

Scriven, M. General Strategies in Evaluation. In Weiss, J. (Ed.), Curriculum Evaluation: Potentiality and Reality, Curriculum Theory Network Monograph Supplement. Toronto: OISE, 1972, 182-192.

Scriven, M. Evaluation Bias and its Control. Paper No. 4 in Occasional Paper Series, Kalamazoo, Michigan: Evaluation Centre, Western Michigan University, 1975.

Stake, R.E. Generalizability of program evaluation: the need for limits. In Educational Evaluation: Theory and Practice. EPIE Educational Product Report, 1969, 2, 39-40.

Stake, R.E. Program Evaluation, Particularly Responsive Evaluation. Paper No. 5 in Occasional Paper Series, Kalamazoo, Michigan: Evaluation Centre, Western Michigan University, 1975.

Stufflebeam, D.L. Meta-Evaluation. Paper No. 3 in Occasional Paper Series, Kalamazoo, Michigan: Evaluation Centre, Western Michigan University, 1975.

# CURRICULUM EVALUATION AND OBJECTIVITY:
# PROBLEMS IN POLICY-ORIENTED RESEARCH

Peter Sheldrake

September 1977

Paper given at
Australian Association for Research in Education
Annual Conference
Canberra November 11-13, 1977

Educational Research Unit
Flinders University of
South Australia

## Introduction

Like every discipline education has its fashions and foibles. In recent years the growing area for researchers has been formative evaluation, and now the new and popular variety is 'policy-oriented' research. Indeed, as a sometime exponent of formative evaluation, I must be careful now in advocating policy-oriented research - there is nothing worse in the fashionable areas of research than to be accused of being fashionable. However, I do not want to appear as an advocate for the policy-oriented approach, as there are more than enough of us offering such advocacy at the present time, especially in this city of Canberra. Rather, my task is to elucidate some of the problems and pitfalls that await the researcher adopting this approach, given our present inadequate understanding of what policy-oriented research is really about.

As is usual in these matters, I want to begin with a piece of clarification: this entails spelling out what policy-oriented research is about. In traditional terms 'research' is concerned with the scientifically based study of phenomena and the relationships among them. 'Policy analysis' is concerned with the specification of policy alternatives, the assessment of alternative means to achieve specified policy intentions, and the identification of the information required to aid such specification and assessment. Ideally, research is disinterested (though certainly not value-free), whereas policy analysis is directed towards various intentions (which are usually explicit and recognised). However, policy-makers require both policy analysis and research to aid them, especially as research is the necessary means by which policy analysis is both developed and improved and, as a result we have seen the emergence of 'policy-oriented research' which is research intended to serve policy analysis.

In practice, three types of research activity have acquired the label of 'policy-oriented'. The first, and probably the most common, is that where an academic researcher goes out and does some 'relevant' research. She or he decides what policy analysts and policy-makers need to know (and may even ask them about this), and then collects data and analyses it in the manner that they (the researcher) feels is most appropriate. Such would-be policy-oriented research is not my interest here: it is actually ordinary (and often good) academic research dressed up to look relevant to practice, but it is accidental if it achieves that end. I do not need to give examples of this.

The second type, at the other extreme perhaps, is research commissioned by policy-makers where they write the research brief and determine the questions to be asked. This is contract research, in my view, and again will not be pursued in further detail in this discussion. However, on passing comment is appropriate, which is that the researchers are often unaware of the 'fine print' of their contracts in these situations, and often have only the request to collect some "neutral, objective, hard evidence", and not the form and basis on which these data were to be collected - Williams (1971) is most perceptive on this.

Finally, there is the research that falls between these two, carried out in the light of policy interests, and in collaboration with policy analysts but not actually directed by those concerned with policy making or analysis. It is to this third, and far more delicate area that I want to direct attention.

## Policy-Oriented Research in Course Evaluation

For some time now we have been familiar with a distinction between 'summative' and 'formative' evaluation, and I think that Parlett and Hamilton (1972) and Stake (1976) have been particularly helpful in clarifying differences here. In the traditional summative approach the questions asked of the evaluator was:

"Is approach X better than approach Y?"

(given various criteria that are agreed and can be used to define what 'better than' means). Such questions demanded a scientific experimental approach to curriculum evaluation, the use of all kinds of sophisticated double-blind research methodologies, and resulted - sadly - in a disappointingly unclear range of results. The same can be said of the related variant of this approved, evaluation based on 'behavioural objectives'.

As a consequence, and partly through a retreat in the face of the complexity of the teaching situation and the impossibility of 'controlling' the many variables involved, a more naturalistic descriptive approach developed - an approach that some researchers in curriculum evaluation see as representing a social anthropological approach: The question now being asked of the evaluator becomes:

"Am I doing a good job?"

a question the evaluator then turns back on the questioner, and in relation to which he tries to 'illuminate' some of the events taking place in preparing and teaching a course.

I now detect a new kind of question in the air, and one that I suspect is going to be asked more and more in the next few years, and the form of that question is:

"Are they doing a good job?"

ahd it is at this point that the policy-oriented research approach becomes relevant to the curriculum evaluator. Such a form of question cuts across the summative - formative continuum. It might appear to be answerable by the strategy of showing that A is doing a 'better' job than B, given the evaluative criteria that the questioner provides, or agrees are appropriate. Equally it could be answered in the illuminative mode, by bringing out the expectations, attitudes and processes that are involved in a given course, and demonstrating the ways in which those involved in a course are successfully managing a mutually acceptable programme: in other words, they are putting together a 'useful' or 'valuable' social construction of reality in the context of the course under consideration.

The point here is that these represent tactics rather like those sometime labelled policy-oriented strategies I outlined before - the first is an example of contract research (if research is the right word) where the problems for the researcher are merely to give operational form to the agreed measures of performance; and the second is an example of research made to appear relevant to the questioner, by telling him or her what they "really wanted to ask". In some contexts one or other of these strategies might be appropriate; in others where the questioner is a policy-maker, the need is for policy-oriented research that feeds back in to the policy-maker's concerns both to develop policy analysis and to take decisions. Such research cannot be conducted by going out and applying agreed criteria (because the criteria are what the policy analyst needs to see developed), nor can it be conducted by describing what is going on (because criteria are needed).

The work of MacDonald and his colleagues on the Nuffield Foundation/ Schools Council Humanities Curriculum Project represents a most important stage in the development of a policy-oriented approach in curriculum evaluation (see MacDonald & Walker, 1976; Stenhouse, 1975; MacDonald in Tawney, 1975), although this project is only one of a number undertaken in the U.K. in the past ten years that have been centrally involved in debates over the interface between research and policy in the curriculum. The University of East Anglia's SAFARI project, Success and Failure and Recent Innovation, edited by MacDonald, Walker, Stenhouse, together with contributions from Elliott, Simons, Stake and Brugelman, has produced an invaluable documentation of this area, (SAFARI, 1974). It is not appropriate in a brief paper like this one to rehearse the sophisticated analysis of the Humanities Curriculum Project that has already been made and MacDonald and Walker's recent book 'Changing the Curriculum' (1976) is an excellent introduction. Rather I want to emphasise the way in which a project originally intended to offer a detailed study of 'how is a teacher in a democracy to handle controversial value issues?' soon found itself in the centre of partisan debate

> "The team saw themselves as a 'research' group offering schools new knowledge and hypotheses about enquiry teaching, rather than prescribing a curriculum. 'We have nothing to recommend,' said Stenhouse. But teacher 'neutrality' was a provocative concept, and they were soon caught up in an inflationary spiral of rhetorical debate, at its most public when their collection of materials on Race was suppressed by the Schools Council in a blaze of publicity. Condemned by N.U.T. and N.A.S. spokesmen, cut adrift by the Council (though not by Nuffield), and assailed variously from the Left ('bourgeois indoctrination'), from the Right ('dangerous revolution'), by academics ('ethical relativism'), and by activists ('substituting social action with a parlour game'), the project was vigorously defended by an equally diverse range of allies, and acquired something of a 'cult' reputation while continuing to compete successfully in the market place."

In a similar way, I found myself in a study of curricular innovations in a medical school in the centre of political debate (reported in Sheldrake and Berry, 1975). At the time I claimed that such debate demonstrated the extent to which I had penetrated the situation under study, and was addressing some rather fundamental reality. I would still claim this was the case, although there are many other problems about that study that I would now, in later wisdom, regret.

## Practice and Problems

To identify an area, and to indicate there are problems, however, is one thing, to analyse these is another. I now want to turn to what curriculum researchers do, particularly in the policy-oriented context, and try to identify the nature of our practice, and the pitfalls that await us. I am going to do so by taking four aspects of the research process - not the intellectual process, but the social process - and these are (1) the initial contract

(2) the establishment of definitions and boundaries

(3) styles, strategies, and ethics

(4) the aftermath

Many curriculum evaluators have stressed the importance of the initial contract - making clear who is going to do what, for what reason, and with what outcomes in mind (cf. Hamilton, 1976; Stake, 1976). In an ideal world this would be done, and would obviate many of the problems that subsequently arise in evaluation exercises (see MacDonald, 1970, as one example among many of this). However, even in a less than ideal world there are some things that can be achieved in establishing the initital contract for an evaluation study, and in the case of a policy-oriented research approach some are crucial.

The central issue in the contract is, of course, control: control here refers both to control over the work itself, and also to control over those doing the work and involved in it. The internal problem is less simple than might first appear to be the case: In traditional 'contract' evaluation the evaluator carries out a study in terms set by the client - there is freedom, to a limited extent over style and strategy, but the basic questions are established for the evaluator, and not by him. In traditional research the researcher dictates both the form of the overall questions and matters of style and strategy.

However, since the emergence of the 'illuminative evaluation' approach, there has been questioning about the extent to which evaluation is necessarily judgemental. Some do not feel that evaluation is necessarily concerned with whether or not something is 'good' or 'bad' (e.g. Parlett and Hamilton, 1972), others do not share this view and see judgement as the core of evaluation (cf. Stake, 1976). For those who prefer the non-judgemental style the issue of control is a delicate one, as they have to manage the expectations of both their clients and themselves over the salience and significance of questions and by virtue of the approach being used, these cannot easily be identified at the outset. As researchers like Parlett and MacDonald have indicated, continued negotiation is essential.

In the case of policy-oriented research it is most important to remember that the purpose of the research is to inform policy-analysis, and so this must be an example where those basic questions that constitute the definition of the research area are determined by the policy-makers, or at the very least in consultation with them. It is only in the areas of style and strategy that control is left more or less firmly in the hands of the researcher. As in the case of illuminative evaluation, this is a delicate balance to strike, though for rather different reasons. On the one hand the policy-makers for whom the research is intended as a means of clarification need to become engaged in the task of defining the areas for study; while on the other hand the researcher has to let go of some of his or her cherished controls on what is going to be done. In emphasising the importance of the distribution of control of course, it is evident that this is bound up with recognition of the importance of purpose, and these two closely relate with one another.

As this discussion runs the danger of becoming somewhat esoteric, an example may help to clarify the issues at stake here. The example to be used is by no means an ideal one, as it is one that relates to a study initiated by myself in collaboration with some colleagues, although from the outset it was deliberately intended to be a policy-oriented study. Rather, its choice is determined by the fact that it is a study that is in the process of being completed at the present time and conducted by the author, and it is full of illuminative errors which can serve as paradigmatic examples of successful strategy.

The study to be used as an example is a project concerned with examining current characteristics and future trends in medical education in Australia, and is being carried out by a project team of five. Before the study began there was considerable discussion of exactly what the purpose and audience of the research was to be. As a primary target the research was intended to contribute to the understanding that policy-makers in medical schools have of their situation and where this situation is likely to develop in the next 10 years: as a policy-oriented study the research was intended to serve their interests.

However, in the initial stages too little attention was paid to actually discussing the nature of the research with those it was intended to help, and it was assumed by the research workers that the intentions of the project were obvious. As with other studies where the contract was assumed rather than made explicit, there was great potential for misunderstanding. In the later stages of the project it became evident that some of our target policy-makers felt that the study was for the researchers' benefit, not their own - and wondered why the researchers were continually anxious to review the research, the data and the observations made. These problems stemmed from a poorly articulated initial contract and were exacerbated by the fact that at least one of the researchers (the author) was as anxious to learn about the methodology of policy-oriented research as he was to identify trends in medical education.

To return to the more general theme of this paper, the main issues that we have so far examined in relation to the initial contract have been control and purpose in relation to the project, and the need to make the contract explicit. Within this general context there are a number of other issues that deserve attention (Stake, 1976 has a further discussion of these in the ordinary evaluation context).

These are, including purpose again for the sake of completeness:

Purpose - Why is the study being conducted?
Specificity - What is it a study of? What are the issues?
Audience - Who is the study for? Who requested it?
Method - What resources are available? What data is required?
Confidentiality - Who will keep the data? What forms of access are being allowed?

As has already been noted, some of these questions are answered in advance in the policy-oriented research context. This includes both those of purpose (though this may need clarification - and often does) and audience. Of the remaining three issues, the last is the most important in the initial contract, in that researchers often seek to use data collected in research for reasons of publication that may not, in fact, be available to them for that purpose. In policy-oriented research the policy-makers or policy analysts must have the right to determine the availability of the data, and not the researchers: their role is to work within the guidelines established and ensure that the normal procedures over the confidentiality of data and the ethics of social research are maintained.

However, even in this area, and certainly in relation to specificity and method, continuing negotiation is both likely and valuable. In the study of medical education introduced earlier one of the principal data collection strategies was a series of interviews with key staff in all the medical schools, intended as a technique for identifying their major areas of concern and interest. In other words this was a procedure for clarifying the issues and hence research questions (and thus relates to specificity) and part of the identification of data itself (method). These data were organised by writing up reports on each 'site visit' to a medical school, and were subsequently reviewed with staff from the medical school. In the course of this review it became apparent that the policy group (the Deans of the medical schools in the first instance) felt it would be valuable to circulate copies of each report to every School - although initially it had been made clear that the data being collected was being kept confidential to each school, and was to be used as a springboard for our own deliberations and investigation. We then had to negotiate this change in method and in the rules over confidentiality with each school, and the preparation of a volume of collected site reports became a second major task of the project.

This discussion of the changing role of the site reports in the medical education study leads me on to the second stage (or process) of the research situation, and that is the establishment of role definitions and boundaries.

115

In the traditional evaluation situation the evaluator was seen as the expert applying various techniques to answer judgemental questions; in the more illuminative tradition the evaluator has tried to evolve a more collaborative, participative role, sometimes remaining detached, and sometimes deliberately taking on an 'interventionist' style (cf. Sheldrake and Berry, 1975). Within the policy-oriented mould the determination of roles is also crucial, and the 'evaluator' (if that is still the word to use) needs to maintain a balance between responding to the requests and interests of the policy-makers on the one hand, and the demands and expectations of the researchers on the other. This balance can best be exemplified in the three areas of objectivity, role boundaries and ethics.

In the case of objectivity, the researcher finds himself urged to be 'detached, neutral and free', and may well aspire to this position himself. However, by virtue of the very nature of the policy-oriented mode, such an aspiration is unattainable - the researcher is being asked to carry out research in relation to, and in order to clarify, policy issues, and they of necessity are not value free. In other words, the researcher is exploring issues in relation to a set of preferred or valued intentions, and while the task is to examine these, or data relevant to these, in a detached fashion, the form of the question is determined by the orientation the research is given. Policy-oriented research is interested research.

In this situation the problems that are most apparent are those of maintaining the research and policy analysis roles for the researcher and his client. For the researcher there is the ever-tempting opportunity to take up policy positions that reflect his own interests and persuasions: in our study of medical education the accusation was made that at times the researchers were working with a 'hidden agenda', (for example to show the value or importance of medical education), and as a result were reminded of the necessary role-boundaries that had to be maintained in this kind of exercise. On the other hand, the client may try to demand characteristics of the research that lead him to stop outside his role as a policy analyst. By telling a researcher that he is not being objective, for example, the client slips from an analytic to a judgemental role about the issues being examined - though that is not to deny that researchers may, at times, fail to meet the requirements of their role!

Finally, reference should be made to ethics. This is, I would argue a particularly delicate area as it is one where negotiated roles and procedures may well fall outside acceptable general definitions of ethical procedure: clients and researchers can readily agree on an appropriate strategy or position to adopt, in view of their shared interests, that unfortunately for them is not an ethical one. At the present time it seems that educational researchers have yet to develop the same kind of rules of ethical procedure that are the case in a subject like psychology - or at least we have yet to make these as explicit and well recognised. However, rules and procedures are only part of the issue, as researchers like MacDonald and Richardson (1974) have indicated, and the ethical problems of the research process are complex and real.

Two features of the research process remain, and comments on these will be somewhat briefer. In part this reflects the extent to which these raise rather less critical problems for the policy-oriented researcher; in part this also reflects the crucial nature of the negotiations over contract and boundaries - from these, much of the rest follows. In the remainder of this paper, then, the areas of 'research styles' and 'aftermath' will be reviewed.

There are, of course, so many research 'styles' now available to us that it is difficult to nominate a few specific problems that deserve special attention. However, a broad distinction can be drawn between three types of research approach. These are:

(1) empirical data-collection approaches

(2) analytic approaches

(3) interventionist data-generating approaches

In the first type falls the more traditional styles of research, the use of tests, survey instruments, panel review, experimental studies, unobtrusive measures, and the like. These need little further comment. In the second category falls research styles that are not concerned with the collection of data as such, but with the analysis of problems in the light of theory and previous research. Again, this approach needs little further comment - in some cases research is carried out almost entirely in the second mode (especially where the data are 'known'), and in others the two approaches are used simultaneously or alternatively.

The third category, of interventionist approaches, does need further comment, however. Research may be conducted by the researcher actually intervening in the research situation, carrying out 'action research' - an approach that derives from Whyte's work in the restaurant and hotel industry in the 1940's, Whyte; (1948) - or by engaging in the 'adversary' mode (cf. Owens, 1973). This is not the place to debate the merits of these approaches, but it is appropriate to signal the consequences that may flow from such techniques where the researcher directly influences the situation that is under study. Where such strategies are proposed they must be evident and drawn to the attention of the client, and their effects on the kinds of research output possibly recognised. However, it is easy to place too much emphasis on this matter for, as others have shown, rigorous traditional research has its unintended implications too (cf. Argyris, 1968).

Finally, there is the aftermath in policy-oriented research. Here, briefly, I want to emphasise one area in particular--the researchers' expectations. There are often problems in research closure, and these are not special to policy-oriented research. However, the fact that policy-oriented research has as its primary client the policy-maker or policy analyst means that the results of the work often may see limited circulation. In an environment where research is rewarded 'invisible' results are a disadvantage and researchers often feel the need to publish something. The danger lies in publishing material that is restricted; the strategy is to negotiate, at the beginning, what the possibilities are, and then accept these.

## Conclusion

Policy-oriented research in education is a new and attractive field. It gives researchers a chance to participate in the world of decision-taking - or at least it provides that illusion. This paper is an attempt to think through some of the problems that such a research style involves, and to develop these in the light of the studies of this kind that have already been undertaken. The most important point to emphasise, I believe, is that participation in decision-taking is an illusion - that policy-oriented research is research for a client, research that may be ignored, that may never be published, and yet that may have its effect on the world. For the realistic, despite problems and limits, it is an exciting mode to work in; for the idealistic it is a road to despondency and dissatisfaction.

## BIBLIOGRAPHY

Argyris, C. (1968): Some Unintended Consequences of Rigorous Research, Psychological Bulletin, 1968 70.3 185-197

Hamilton, D. (1976): Curriculum Evaluation, Sydney: MacMillan

MacDonald, B. (1970): Evaluation of the Humanities Curriculum Project: A Wholistic Approach, Mimeo: University of East Anglia

MacDonald, B. (1975): Chapter on the Humanities Curriculum Project, in D Tawney (Ed), Curriculum Evaluation Today: Trends and Implications, London: MacMillan

MacDonald, B. & Walker, R. (1976): Changing the Curriculum, Sydney: MacMillan

Owens, T.R. (1973): Educational Evaluation by Adversary Proceedings, in E. House (Ed), School Evaluation: The Politics and Process, Berkley: McCutchan

Parlett, M. & Hamilton, D. (1972): Illuminative Evaluation: A New Approach to the Study of Innovatory Programmes, Edinburgh: Edinburgh University, Centre for Research in the Educational Sciences, Occasional Paper 9

Richardson, E. (1974): The Teacher, The School & The Task of Management, London: Heinemann

SAFARI (1974): Innovation, Evaluation, Research & the Problem of Control: Some Interim Papers, Norwich: University of East Anglia, Centre for Applied Research in Education

Sheldrake, P. & Berry, S. (1975): Looking at Innovation, Windsor: National Foundation for Educational Research

Stake, R.E. (1976): Evaluating Educational Programmes, Paris: Organisation for Economic Cooperation and Development, Centre for Educational Research & Innovation

Stenhouse, L. (1975): An Introduction to Curriculum Research and Development, London: Heinemann

BIBLIOGRAPHY (continued)

Whyte, W.F. (1948): Human Relations in the Restaurant Industry, New York: McGraw-Hill

Williams, W. (1971): Social Policy Research and Analysis, New York: Elsevier

# Curriculum Evaluation and The Development of The Person

F. John Hunt

Monash University

Paper to be presented at 1977 Annual Conference of the Australian Association for Research in Education, Canberra 11-13 November.

# Curriculum Evaluation and the Development of the Person

F.J. Hunt

The task of curriculum evaluation has become of considerable significance with the rise in the social significance of schooling during the 1960s and '70s. But, as with so many areas of social activity, much effort has been concentrated on technical aspects of tasks, with a corresponding neglect of the social processes in which the techniques are employed. A major purpose of this discussion, therefore, is to examine curriculum evaluation as a social process, giving particular attention to patterns of participation and to implications of different patterns. Now when we do that, we readily recognise that curriculum evaluation, as with other social processes, reflects underlying assumptions about the nature of people and their interrelationships, including the relationships between specialists such as evaluators and the people with whom they work. Hence a second major purpose is to explore the relationships between values and ways of participating in social processes, with specific reference to curriculum evaluation. We can get at these relationships by examining some alternative conceptions and then noting similarities between particular conceptions.

## Approaches to Curriculum Evaluation

Although curriculum evaluation is the topic for central consideration, it is not the most useful point from which to start. Indeed, we cannot go very far until we decide what we regard as curriculum, for it is that which we are concerned to evaluate.

Now the definition of curriculum, as with other social phenomena, is particularly arbitrary. For one thing, a definition can only be expressed in very general terms, simply because curriculum as a phenomenon is differently conceived by different people. That is more or less inevitable: given that people have different biographies, interests, aspirations and expectations, they can be expected to perceive and conceptualise particular phenomena in different ways. Even where people are similar in those respects, they may still vary greatly in what they include as part of a particular phenomenon.

In respect of curriculum, to return to our concern here, some have found it appropriate to include only the planned materials and courses of action, such as might be included in a particular project or set of units.[1] To do that, however, is to restrict curriculum to only part of the situations and events which children experience and are affected by. In particular, it excludes the unplanned interactions and contributions of teachers and children and possibly others. It also excludes experiences which flow from the ways schools are organised and administered, such as arrangements of being zoned to a particular school and assigned to a particular class, teacher and program. It excludes the situation and the process where administrators make regulations that determine rights and responsibilities that teachers, parents, children and possibly others may or may not have, and thereby the character of their relationships. It excludes, too, the financial systems which commonly function to take resources from people and redistribute them in some way in the course of offering parents and children a limited range of options in schooling. Each of these, and others, is a means for influencing participation and learning in the schooling process. In that they bear upon children and influence their learning and development, they can be included within the definition of curriculum which is then seen as very comprehensive.[2]

A second point for diversity in positions arises in relation to how curriculum evaluation is conceived as a process. One tendency is to limit and separate, and to define very specifically: thus Cooper sees curriculum evaluation as '*the collection and provision of evidence*, on the basis of which decisions can be taken about the feasibility, effectiveness and educational value of curricula'[3] (emphasis added). In contrast with that is a view that sees evaluation as integral with other processes in the general process of schooling. Thus we can identify and include the determining of aims and objectives, the gathering together of resources and the determining of means to be employed, the undertaking of schooling activity, evaluation of the functioning of that activity, the analysis and interpretation of results, the dissemination of what might be termed findings and conclusions, and deciding what to do next. In that these constitute a series of steps or processes which bear upon the schooling experiences of students, they can be seen as part of the curriculum and constitute areas of expertise in which students need to develop competence if they are to become independent, autonomous learners. That is, curriculum evaluation can be seen as a separate and specialised activity or as an integral part of the curriculum and thus of the schooling experience.

We can now move to the third point of participation to which I want to give special attention. As previously indicated, I propose to focus on the social aspect, considering who participates and in what ways, rather than on technical matters. In considering that issue it can be useful to conceive two basic polar or ideal-type models or modes of participating in curriculum evaluation to which we can relate discussion.[4]

One is a situation where only a relatively few people participate in any meaningful sense, and then in a process that is narrowly defined, that is, different people or sets of people undertake each step or process. In such a mode of operating, participants in each step commonly comprise specialists with substantial expertise who act in ways and with techniques that may not be widely understood. Correspondingly, there can be considerable mystification and even alienation among non-specialists.

An extreme contrast with that would be a mode of operating in which all interested parties participate in most if not all of the several steps involved in schooling. Thus, teachers, children and parents might participate in determining aims or objectives, in determining what resources and means will be employed, in undertaking the teaching-learning activity, in appraising its functioning, in drawing conclusions from their evidence, and in making decisions about what will follow. In such a mode of operating a specialist evaluator, as an administrator, curriculum developer or other type of specialist, would act essentially as a resource person, acting in a service capacity, enabling teachers and children and parents to act according to their judgements.

Now a particular point I want to make is that the extreme positions on these several issues are interrelated and cumulatively support particular, more general modes of operating. To limit and exclude in respect of curriculum evaluation - of what counts as curriculum, of what is involved in evaluation, and of who participates and in what ways, is to define, control and direct, to foster lateral and hierarchical specialisation with limited rights and responsibilities. It is to foster the process of bureaucratic administration which, along with such other processes as positivistic-behaviouralistic social science, assembly line production systems and much of modern technology, effectively serves central or élite control and management. On the other hand, openness and inclusiveness in what counts as curriculum, in what is involved in evaluation and in who participates and in what ways, entails a more flexible, flatter social structure, a dispersal of power, including expertise and competence, and facilitates such interactive processes as dialogue, negotiation, compromise and shared responsibility. Particularly when supported by appropriate social science, modes of production and technology, it can favour a more actively participatory style of decision making and organisation.

It will also be obvious that we do not choose between alternatives such as these simply on technical grounds. Fundamental to any choosing will be the values upheld: whether, to polarise further, we prefer individual development with its inherent diversity, with active participation of each person in determining events and making decisions, and widespread rights and opportunities to experience achievement, fulfilment and satisfaction or, alternatively, greater emphasis on the transmission or inculcation of particular cultural models and lifestyles, the control and moulding of less powerful people, with passive accommodative responses and life styles for many people. Influencing our choices concerning styles of participation, for the specialists as well as for the lay person, will be preferences for types of people and patterns of relationships.

Now this is not to argue that each model works in precisely the ways indicated here. For one thing, human beings are remarkably resilient as well as accommodative, able to avoid the restraints and pressures of closed institutions as well as resist the facilitative scope of more open ones. Nonetheless, the logic of each model is in a particular direction so that the social forms can influence the kind of development that occurs.

Nor is it to argue that we should totally reject one model or another. Rather, given the general pattern of human development, a useful accommodation can be to see the two not as discrete alternatives but as extreme points on a continuum between which people can move. Because human beings commence life as very dependent people there is an initial need for adults to make decisions and act on behalf of infants. But given that a goal is the development of the person, of increasing one's competence to handle situations and events, then the goal in any process, be that determining what we will learn about, how one will go about learning and appraising one's experiences, and making subsequent decisions in the light of that appraisal, should include the increasing participation in those processes of the hitherto more dependent people. Thus, initially in respect of schooling, decisions might be made mainly by parents and teachers, in a context of supportive activity by administrators and others. But progressively children should take a more active and substantial part in those processes as they move towards autonomy and responsibility.[5]

## An Evaluative Framework

We are now in a position to draw upon the main themes considered here as the underpinning for an evaluative stance, and formulate some specific questions to be asked in evaluating curriculum. In doing that, I do not imply that the stance adopted is a comprehensive one that might replace other evaluative frameworks or positions such as the Curriculum Materials Analysis System described by Knight and Hodges[6] or ones developed for studies of specific projects.[7] Clearly, it is constructed to take concern with a particular although central facet of curriculum, namely the conception of people and their rights and obligations - that is, with their relationships with each other, and thereby the nature of the social entity of which they are part, be that a group, community, association, organisation or society. The concern taken up here is more akin to the ideas expressed by Toms in his identification of a preoccupation with types of people exemplified in curriculum.[8] In practice, it could be a part, albeit a central part, of any framework utilised in an evaluative study of what is happening through the medium of curriculum.

Of the questions that can be conceived as central, one concerning assumptions about people and their relationships can usefully be asked in terms of whether people actively or passively participate in situations and processes. A second can be asked about how different groups or categories of people are represented. Thus we might ask of means to be employed, be they material resources, teaching strategies or organisational arrangements:

- Are people represented as relatively active or passive participants in situations and relationships?
- What distinctions are made between people in relation to such physical or socio-cultural characteristics as ethnicity, colour, sex or class?
- Again, what distinctions are made between teachers, parents and children, and between them and various types of specialists?

A similar set of questions might be asked of ongoing activities. In particular, we might ask who participates, and in what ways:

- In determining aims and objectives?
- In determining the materials to be used?
- In determining the means to be employed?
- In undertaking the teaching-learning activity?
- In evaluating what has been achieved?
- In deciding what is to be done subsequently?

Again it is also pertinent to ask

- Are distinctions made between teachers, parents and children?
- Are distinctions made between different types of people in relation to such physical or socio-cultural characteristics as ethnicity, colour or sex?

I recognise that these questions are crudely expressed and that we may want to know much more about participation in particular areas or at certain points within particular processes. Because my primary concern is to illustrate an overall approach, I want to leave the questions and the stance at an elementary stage for the present, and attempt to sort out a few of the more general issues. We can do this to some extent by applying the perspective and some major questions to a few well-known curriculum programs.[9] As one, I have chosen the Australian Science Education Project, developed by the Australian governments in cooperation between 1964 and 1974. As a second, I was interested in examining the Social Education Materials Project, the recently completed (but not yet published) social education project also undertaken by the Australian governments in cooperation. Because it was unavailable I selected instead Australian History, the subject with the third highest enrolments after English - a compulsory subject - and Biology in the Victorian High School Certificate program which is located at the end of the sixth form of secondary schooling. As a third, I have examined the Humanities Project developed by Stenhouse and his colleagues for the Schools Council in Great Britain. Variations in the character of these programs make them of particular interest for trying out the evaluative framework argued for in this paper.

Australian Science Education Project.[10] The Australian Science Education Project (ASEP) was the first curriculum development project undertaken on a national basis in Australia, being funded as a joint effort by the Federal and State governments over the period from October 1969 to March 1974. The Project was seen by its sponsors, acting on advice from their curriculum authorities, as producing 'a range of instructional materials sufficient in quantity to satisfy a major proportion of the requirements for courses in secondary school science for Grades 7 - 10', that allowed for individual differences, particularly in respect of stages of intellectual development, interests, reading levels and rates of working, took account of similarities and differences in science education in all states, and accepted that the teacher should have final responsibility in deciding what will be done in his or her classroom. Incidentally, it was seen to offer 'exciting possibilities' in having the resources to make a national contribution by planning materials from which teachers could choose when planning courses and by demonstrating one effective method of curriculum development.

The general broad aim of ASEP might be regarded as student focussed, with a science bias - to design science experiences which would contribute to the development of children. The science component was derived from a set of ideas which became known as the Environment Scheme. This had five main parts comprising:

- the ways in which human beings resemble and differ from other individual organisms
- how interaction among groups affect decisions made by man
- the ways in which man has extended his ability to explore and utilise his environment
- the ways in which technology has changed man's environment, and
- the changes in the environment which take place naturally, and the ways in which man has interfered with these changes.

It was believed that experiences in these areas could develop

- some understanding of man, his physical and biological environment, and his interpersonal relationships
- skills and attitudes important for scientific investigations
- some understanding of the nature, scope and limitations of science
- some understanding of, and concern for, the consequences of science and technology.

It was incidentally believed that the kind of understanding at which the Project aimed would enable children to operate more effectively in their environment, and that the arousal and fostering of interest was of prime importance in the development of understanding, skills and attitudes.

When we turn to the material produced by ASEP we can readily recognise that a number of additional assumptions influenced the character of the Project's productions. Some relate to knowledge and are indeed conventional assumptions of scientists:

- that there is a world 'out there' and it is knowable
- that there are ways of finding out about that world
- that some aspects of that world are more important or of more interest than others.

A second set of assumptions relate to the attributes and activities of people.

First, there are specialists who know a great deal about the world or particular aspects of it, and these people can help teachers to help children learn about the world.

Second, teachers are able to organise materials and experience, to motivate children, coordinate their activities, appraise their work, and generally guide their learning activities. For such reasons they are valuable advisors in the development of units and materials.

Third, children are seen to be inherently active and (largely following the ideas of Piaget and other developmental psychologists) to maturate and develop through experience. Conceivably their activity can be utilised in interaction with science materials so that they are affected by or respond in ways that influence cognitive, affective and motor development. Of significance too, are assumptions that children observe, compute, infer, and have interests which influence what they study. This justifies offering sets of options from which children can choose.

Fourth, parents are recognised as significant to the child's learning and development so that their understanding and support are matters of concern. It is also recognised that some material bears upon beliefs shared within a family and may not be acceptable to some parents. Thus, particular units or topics are recognised as sensitive and parents' reaction are to be taken into account.

The situation and the assumptions indicate a point of tension that pervades the schooling process. On the one hand, active dynamic children are responding to their own needs, interests and concerns, and to stimuli from the world at large. At the same time, teachers have the task of confronting them as students with selected experiences so that their development is influenced along certain lines - towards certain areas of knowledge, understanding, skills and attitudes. That is, schooling constitutes an attempt to control and direct development, with important ramifications for the development of the individual and the functioning of society. Again, concentration on control and direction towards certain ends means, commonly although not necessarily, the preclusion of other possibilities, although presumably if such deficiencies or omissions are recognised they can be remedied.

Sometimes this predicament gives rise to consequences of a very serious nature, as is illustrated in the unit on Australians - Past and Present. In that unit Aborigines are perceived as interesting phenomena, with questions taken up concerning theories of their origins, their distinctive characteristics and group ways, and with archaeology and anthropology seen as ways of learning about them. Live Aborigines, confronted with problems of surviving in contemporary society, may well feel that being considered an interesting specimen is unsatisfactory and, more seriously, misses the main point of their situation and circumstances in the present. Conceivably, important science experiences could still have been obtained in the course of considering Aborigines as live and vital people who might be interested in scientifically studying issues and problems currently confronting them.

Some partiality or bias in perception is reflected at a number of points in units in respect of other groups of people. In general, girls are shown as similar to boys in the activities they engage in, but nevertheless there is still some tendency for sex role differentiation in the illustrations showing tasks that people undertake. Again, there is a neglect in these examples of the ethnic pluralism that characterises our society. To some extent this may reflect a preoccupation with the traditional concerns of scientists - but that is to pass over factors which have generated those concerns. This observation draws attention to a further point that the ASEP material does not include an examination of relations between social factors and scientific activity - what has come to be termed the sociology and politics of science. Conceivably, such a topic may be inappropriate for the selected population of junior secondary school students, but again something might have been attempted. In any case these several examples, taken together, give support to an interpretation that the science considered is the science of white anglo-saxon and possibly male Australians and does not direct attention to how science as a way of responding to particular problems may be made relevant to other cultures and life styles.

At one level then children and teachers are seen as active, and that activeness is taken into account with suggestions on ways in which they might participate. But when we look to the more comprehensive set of processes that make up schooling - to the determination of aims and objectives, material and experiences and organisational formats to be used, to the undertaking of the teaching-learning activity, the evaluation of what has been achieved, and to deciding future courses of action - we find their participation is very limited. Moreover, the pattern is basically the same in each unit so that no expansion of student or teacher participation in decision making is indicated. Thus the model for roles and participation, for the development of competence in the various processes associated with schooling is a static one: children are not led to progress from novitiates to some degree of competence in such activities as determining aims and objectives or deciding future courses of action. At the same time, teachers, seen to have responsibility for their programs and involved in the teaching of units, have opportunities to participate and so a definite and more or less unchanging distinction is made between teachers and students.

Certain explanatory factors can readily be identified. One undoubtedly, is the set of assumptions that reality exists 'out there', that knowledge about it is discoverable and transmissable, that certain specialists know about the world and have developed techniques for learning about it. The ends and means of science education are largely framed by those assumptions and act to limit the scope for participation by teachers and particularly by students. A second factor is the concept that the ASEP units should be a set of resources that can be employed by teachers in a variety of ways to supplement many different types of programs. That policy meant the avoidance of sequencing and cumulative development of skills and competencies in decision making and participation in the wider range of processes that comprise schooling. Sequencing is not altogether ignored: on a number of occasions possible interrelationships between units are discussed so that background may be built up for particular units. It is sequencing in the wider sense that is left aside. While that was done largely to avoid an appearance of being overly directive, which is an important consideration, the omission is also an important one, particularly once we consider curriculum and its evaluation in the larger context of schooling.

Australian History:[11] The subject Australian History is taught within the framework of the Victorian Universities and Schools Examinations Board which, as was previously noted, conducts examination programs that serve as a basis for selecting (and screening out) students at the end of secondary schooling for certain areas of employment and tertiary studies. From a student's perspective, the program is a means for establishing one's claims to selection or employment, in competition with other students. The location and functions establish many of the characteristics of the subject - a common program to be taught throughout Victoria, a course that is prepared by one group of people, a standing committee of the Board drawn from Universities, schools and schooling systems; taught by another, the teachers in secondary schools; and examined by a third, a group established by the standing committee. In certain respects - that the course provides a basis for grading students - the particular content is of little consequence. At the same time, in that it is presented in the name of a particular subject, the content is of significance ~~for~~ as an expression of the operationalisation of underlying assumptions and conceptions of what constitutes a course in this area of history.

Teachers of Australian History are first directed towards a general set of objectives that serve for all histories taught in the Higher School Certificate program. These objectives relate to knowledge, responding with historical resources to certain types or forms of problems, the expression of ideas and, throughout, developing an understanding of the mode(s) of operating of historians. Such objectives are not necessarily required by the VUSEB framework but represent an adaptation by historians that enables them to work within such a framework.

The specific content of Australian History involves more particular concerns. One is to structure and focus the history of Australia as what has transpired since the first settlement as a story of national development. This is done in three parts. The first, dealing with the period 1788 to 1850, has as its main theme 'the planting of a penal colony in New South Wales and its transition into a free society'. The second, dealing with the period 1850 to 1901, focusses on 'economic, political and social development in the colonies, especially Victoria and New South Wales after 1850, and on the movement towards nationhood'. The third, dealing with the period 1901 to 1950, is 'concerned with domestic development and foreign relations of the Australian Commonwealth'. Accordingly, the course can be conceived as focussing on national development, and in a way that would seem to give a sense of inevitability to that development. In brief, it is a structured view with defined priorities as to what counts in the history of the society.

Several important implications follow from these and associated features of the course. One is that the course does not, obviously at least, encourage a more idiosyncratic and personally meaningful structuring of Australian history. While undoubtedly the three periods are defensible,[12] they nonetheless carry with them a requirement that teachers forego alternative ways of interpreting the field. Thus they impose points or principles about which there shall be uniformity throughout the state of Victoria. Conceivably, a substantial proportion of students taking the course are likely thereafter to view Australian history in those terms. Correspondingly, teachers and students are thereby assigned a passive accommodative role in the process of developing conceptions of Australia's history.

Further interesting emphases or biasses are apparent in a closer inspection of the details of the course outline.

It is significant that the course begins with the settlement in 1788, including the British background to that event. There is no definite reference to the situations in Aboriginal tribal societies prior to the arrival of the whites so it would appear that the impact on Aboriginal life cannot be fully appreciated. It would also seem reasonable to argue that the settlement is viewed from a particular type of white person's perspective. Perhaps the historian's mode of operating, namely relying largely on written records, predisposes one to the perceptions of those who wrote, mostly middle and upper class people - officials, free settlers, merchants and the like.*

But the bias does not end with the beginning. It continues to be evident as we ask about particular groups in Australian society. We can follow the treatment of Aborigines a little further.

Section 1 (1788 to 1850) of the course is subdivided into four parts - (a) The Early Years, (b) The Colony under MacQuarie, (c) Towards a Free Society and (d) The Squatting Ascendency. Aborigines are referred to in parts (a) in the topic 'European attitudes towards the Aborigines and their culture' and (b) with 'Policies towards the Aborigines'. They are referred to again in part (d) in three topics - 'Conflict between pastoral and Aboriginal use of land; attempts by governments to regulate contact between Europeans and Aborigines; the impact of European contact on Aboriginal society'. Presumably they would be considered in other topics but these are the ones that are highlighted by specific reference. It is also worthy of note that the only other specific reference to Aborigines in the remainder of the course is in Section 11, 1850 to 1901, part (c) - Politics and Policies in Victoria and New South Wales, 1860 to 1900, when the bare statement 'Aboriginal policies' is included. Here again, of course, this would surely be a consideration of governmental policies and not of the efforts and concerns of Aborigines.

We can extend our evidence by examining the list of selected texts included in the handbook. The authors acknowledge that the list is not exhaustive but it does represent a selection from material available at the time of preparing the outline. From the list, three major works dealing specifically with Aborigines are included among the references for part (a) The Early Years, in the first section. They are referred to again in each of the subsequent parts of that section. Two of the three are referred to again in part (c) in the second section on Politics and Policies in Victoria and New South Wales, 1860 to 1900. Finally, three pamphlets are included in a list of material available from the Victorian Historical Association. Clearly, again, of course, Aboriginals are discussed in other more general texts or in discussions focussed on other topics, but these are the only explicit references focussing on them.

---

*At the same time, recent years have been a period of considerable integration, eclecticism or whatever, and a more comprehensive approach should not have been difficult to devise and teach.

On the basis of explicit reference in the course outline and suggested reference reading, a student might be forgiven for concluding that Aborigines disappeared from Australian society, and not simply from Tasmania, sometime in the late nineteenth century. Presumably such a misrepresentation does not derive from conscious prejudice or bias but rather follows from a particular perspective that gives primacy of concern to certain aspects of Australian history, namely its development as a nation, and neglects how particular groups such as Aborigines fared in the emergence of nationhood. That is, the biasses in the perspective emerge systematically in the course. We can test that interpretation further by examining the treatment of other groups.

A second group whom we may consider is women and a perusal of the course outline does not identify a specific reference to the presence of women in Australian society. It could also be argued, of course, that there is not a specific reference to men either, but then the course refers very much to issues in which men have been dominant and determinative. There would surely have been adequate grounds for at least occasionally looking at the separate experiences of men and women. But then, there has not been opportunity to look at the problems and circumstances of other groups such as children, the aged, or the poor, except as these have attracted welfare policies or other responses from the developing nation.

It is tempting to interpret the overall course outline as a product of several major factors. One, a need for the assessment and grading of children, supports an emphasis on a common course, involving facts and some skills with analysis and techniques of historians. A second, a preoccupation with teaching history, limits concern to issues that are reported or examined in the historical literature and results in the neglect of issues not so dealt with. A third, and pervading the whole course, is a preoccupation with the emergence of a national entity with, incidentally, consideration of particular groups only as they impinge on the development of the nation-society. The neglect of women, the aged, the poor and other particular groups is probably a fair indication of their relative lack of influence on the development of Australian society.

But significant, albeit subtly, in all this is the unobtrusive use and, conceivably, transmission of a particular perspective. As noted earlier, it would appear to be the perspective of the more affluent and literate, white, older males, preoccupied with the emergence of a national-social entity. Thus children are steadily inculcated into a perspective - a structuring of the time sequence, a selection of issues, and to modes of examining those issues - and ignoring others. But they are not passive participants only in those respects; they are also passive in all steps of the teaching-learning process. They can but accommodate to the structure and arrangements and hope to be judged more or less meritorious on their accommodative efforts.

The Humanities Project.[13] The Humanities Project, sponsored by the British Schools Council and the Nuffield Foundation, was set up in September, 1967 and ran for five years. The humanities was defined as involving the study of important human issues and the aim of the project was to help with 'the development of an understanding of social situations and human acts and of the controversial value issues which they raise'. Values were seen as basic and, given the plurality of value positions in contemporary societies, certain issues would invariably be controversial. Issues selected for experimental study and for the preparation of resource material were war, education, the family, relations between the sexes, poverty, people and work, living in cities, law and order and race relations. It was anticipated that the project would come up with contributions that would extend the range of choice open to teachers working in the humanities with adolescents of average and below average ability.

The need to come to terms with controversial issues, in the sense of developing a mode of dealing with them, together with a view of schooling as education - 'committed to a preference for rationality rather than irrationality, imaginativeness rather than unimaginativeness, sensitivity rather than insensitivity . . . for respect for persons and readiness to listen to the views of others' - rather than transmission and training, is reflected in five major premises that underlay the project. They are

- That controversial issues should be handled in the classroom with adolescents;
- That the teacher accepts the need to submit his teaching in controversial areas to the criterion of neutrality at this stage of education, i.e. that he regards it as part of his responsibility not to promote his own view;
- That the mode of enquiry in controversial areas should have discussion, rather than instruction, as its core;
- That the discussion should protect divergence of view among participants, rather than attempt to achieve consensus;
- That the teacher as chairman of the discussion should have responsibility for quality and standards in learning.

It is obvious that the project assigned a central and particularly difficult responsibility to the teacher. At the same time as being neutral or impartial, the teacher was also expected to initiate, organise, establish a congenial environment, decide upon resources that might be brought into the discussion, be sensitive to the circumstances of students and to delicate issues in relationships, protect divergence of view, mediate critical standards which support work of quality in the group, and undertake other tasks. In effect, the teacher was to exemplify a version of the educated person.

The project also anticipated active participation by students. It was recognised that individuals bring into group situations such particular attributes as their desires, wishes and feelings. Again, while the chairperson has a responsibility to develop a congenial environment, s/he also has a responsibility to encourage critical appraisal of his or her performance. The students are also seen as likely to be the ones to decide upon a topic, to take initiatives and responsibilities within the discussion, and in extending activities beyond the discussion, such as to obtain further information and possibly undertake a research project. But, centrally and fundamentally, the teacher has a responsibility to teach students a set of procedures for the orderly and fruitful conduct of an enquiry into a controversial issue. To the extent that students accept that framework, they were free and welcome to develop competence, initiative and other dynamic characteristics.

Interestingly, the material is probably less structured and determining of reactions, or requiring particular accommodative behaviour than that of most projects. The sets of material on the several issues comprise print and non-print forms and constitute a wide ranging selection of products of people that bear upon the selected controversial issues. It is offered as a resource to be drawn upon according to judgment as to their relevance and usefulness, and supplemented as opportunity occurs. Its relative unstructuredness is consistent with the position and strategy of according responsibility for decision making to the teacher and the group. It also gives particular significance to the style of operating which is not spelt out in descriptive or other material but which constitutes the traning programs for teachers.

It can be argued that the Humanities Project is one of the bolder ventures in the sense that it gambled for greater stakes and takes greater risks than have most. In selecting controversial issues as its focus it takes concern with a most sensitive aspect of social relationships. In seeking to devise and gain adoption of a mode of handling such issues it attempts to contribute in a fundamental way to holding societies together. By placing reliance on the teacher and seeking the cooperation of students it attempts to bypass common, more authoritarian modes and structures. Not surprisingly, the project's approach itself has been very controversial, alienating some and attracting others with a degree of fervour. MacDonald (who worked with the project) and Walker offer the interpretation 'that Stenhouse invited the teacher to create an alliance with him against the forces of institutional and attitudinal inertia in the school system'.[14]

In short, curriculum was defined in a fuller sense than is usual and, correspondingly, the project has had wider ramifications than are usual. There is room for a wide range of active participants, given that they accept the mode of handling controversial issues. But if they do not accept that mode, it would be because they actively reject it and withdraw from the process or seek to destroy it. In either case, the project, acknowledges the active nature of people - particularly of teachers and students.

**Comment**

A concern with people and their participation in social processes provides a useful and important basis for appraising institutions and their policies and practices. In working from that basis it is further useful to consider two major possibilities - one that might be termed a relatively closed and limiting position and, its extreme contrast, a more open and including position, and to see them as ends of a continuum between which people may move as they increase in competence, autonomy and responsibility.

Schooling, the location of curriculum and curriculum evaluation, provides a particularly crucial situation for the exercise of social influence. In the extreme cases, particular participants can limit the scope for others to participate and channel development along particular dimensions, or enable diverse and idiosyncratically meaningful personal development to occur.

That possibility raises important questions about how specialists participate in a process such as curriculum evaluation, and the kinds of influence which they exercise on teachers, students and others. Given an emphasis on technical matters in our type of society, it has been common to pass over social issues, even to fail to take account of their existence as products of prior socio-political activities and decisions, and thus neglect the influence of those activities and decisions, as well as of the functioning specialists, on the development of people and society. While such taking-for-grantedness is understandable, due to the force of socialisation processes, it nonetheless constitutes a form of socio-political action for which participants have a responsibility to be conscious, so that they are witting rather than unwitting in selecting a mode of participating in a process.

At a ~~[illegible]~~ simple concrete level, in this paper I have used three programs to examine how people are represented in means employed in curriculum and how they act in various processes of which curriculum evaluation might take account. The approach reveals important differences between the three programs in the scope available for participation and, presumably, for the development of competence, autonomy and responsibility. In particular, it makes clear the controlling and limiting nature of courses conducted as part of a system of external examinations that serves as a sifting and sorting mechanism to select for (and screen out from) tertiary education and certain areas of employment. It also identifies and reveals several examples of discrimination in the manner of representing different types of people, so that programs can be seen as racist, sexist or otherwise offensive and even damaging. Such features could be very relevant in explaining the relative usefulness of particular curricular projects.

The illuminativeness of the approach and the illustrative analyses also indicate that much useful work can be done at a fairly general and non-technical level of analysis. Indeed, the point can be made that a preoccupation with technical issues can serve to limit perceptiveness and lead to failure to recognise some of the more serious biasses in social processes. The curious irony is that a preoccupation with technical issues has often been interpreted as indicating an impartial, objective, and even value free approach, while all the while such work passed over some of the major forms of discrimination and restrictive influences that are exercised through social processes.

Even so, it has not been my intent in this paper to argue for the rejection of refined techniques or to uphold Ludditism in some form. Rather, my point has been to argue for a clarification of assumptions and orientations, and to locate techniques within a perspective that comprehends our fundamental purposes and is sensitive to the extensive and pervasive ramifications of our actions.

**References**

1. e.g. Gagné, R.M., 'Curriculum Research and the Promotion of Learning,' in R. Tyler, R. Gagné and M. Scriven, *Perspectives of Curriculum Evaluation*, AERA Monograph, Chicago, Ill., Rand McNally, 1967, quoted in Hugh Sockett, *Designing the Curriculum*, London, Open Books, 1976, p. 23.

2. Kansas Curriculum Guide for Elementary Schools, quoted in Lawrence Stenhouse, *An Introduction to Curriculum Research and Development*, London, Heinemann, 1975, p. 2.

3. Cooper, Keith, 'Curriculum Evaluation - definition and boundaries,' in David Tawney (ed.), *Curriculum Evaluation Today: Trends and Implications*, London, Macmillan, 1976, p. 10.

4. This discussion draws upon a statement of positions that I plan to develop more fully on another occasion.

5. Barry MacDonald has expressed the alternatives in terms of three possibilities - bureaucratic, autocratic and democratic; see his 'Evaluation and the Control of Education,' in Tawney (ed.), *op. cit*. The dependence of the autocratic evaluator on 'the research community, and *high levels in the bureaucracy*' (emphasis added) is recognised by MacDonald, p. 133.

6. Knight, Merle M. and James O. Hodges, 'Curriculum Materials Analysis System: A Summary of Experience,' *SSEC Newsletter*, No. 7, May 1969, pp. 1-5.

7. e.g. as in the evaluations discussed in Eraut, Michael, 'Some recent evaluation studies of curriculum projects - a review,' in Tawney (ed.), *op. cit*.

8. Tomes, Alan, An Approach to Selecting Among Social Studies Curricula, St Ann, Missouri, CEMREL, 1970 (Mimeo).

9. For a more detailed illustration of what emerges when subject material is appraised from such a perspective, see the analysis for sexism in consumer education and mathematics texts included as appendices in *Victorian Committee on Equal Opportunity in Schools*, *Report to the Premier of Victoria*, Melbourne, July 1977 (Mimeo).

10. The discussion that follows is based substantially on the Australian Science Education Project explanatory booklet, *A Guide to ASEP*, Melbourne, Government Printer, 1974, together with booklets on particular units but especially *Australians-Past and Present* and *The Human Machine*.

11. These comments are based on the statement 'Higher School Certificate Australian History,' in *Handbook of Directions and Prescriptions for 1975*, Melbourne, The Victorian Universities and Schools Examinations Board, October 1974 (The last year for which enrolment figures were available at the time of writing).

12. In 1975 the course was restructured from a previous outline of four periods - 1. The Early Colonial Period: From First Settlement to 1850; 2. From Gold to Depression: 1850 to 1893; 3. The Creation of the Commonwealth; and 4. The Commonwealth of Australia Before World War 11. *Handbook of Directions and Prescriptions for 1974*, Melbourne, The Victorian Universities and Schools Examinations Board, October 1973.

13. The following discussion is based on the introductory booklet, *The Humanities Project: an introduction*, London, Heinemann, 1970; the teachers' handbooks for the several topics and the sets of materials.

14. MacDonald, Barry and Rob Walker, *Changing the Curriculum*, London, Open Books, 1976, p. 84.

# The Use of Criterion-Referenced Testing in Curriculum Evaluation

Sid Bourke and George Morgan

## The Core Curriculum

School curricula may provide for a wide variety of student outcomes, some specified, some not specified. In areas such as the fostering of creativity and individual development through the following of personal interests and inclinations, student outcomes are extraordinarily difficult to specify and such specifications as can be arrived at are frequently forced and specious. Some would argue that even if student outcomes can be specified in these areas, they should not be specified because of the constricting effect they would have on essentially open-ended programs.

There are, however, other areas of school curricula where the specification of student outcomes is more straightforward, and indeed far more important. These areas are related to the knowledge and skills that all students should achieve while at school so that they will be able to progress through the school system as independent learners as far as they wish, and so that they will be able to play a full and active role in the wider community on leaving school. The knowledges and skills concerned here would form the core curriculum of a school and, we would argue, should be clearly specified and should be at the centre of every school's program. School curricula would then consist of at least two segments: a common core of specified essential knowledge and skills and an idiosyncratic collection of other activities, some with and some without specified student outcomes. Smith and Orlosky (1975:89-90) refer to these two segments as fundamental learning and advanced learning.

The concept of a common or core curriculum is not new in Australia - indeed it was not long ago that virtually all of the school curriculum, centrally laid down, was treated in this way. However, with more recent developments towards region-based and even school-based curricula there would seem to be a danger of overlooking the importance of essential abilities that all students must possess if they are not to be disadvantaged in progress through the school or in their everyday lives while at school and in later life. At the very least, the essential components of any curriculum are the abilities to read, to write and to work with number - literacy and numeracy are the rights of every student. It has been suggested that the imposition of a common core of subject-matter content in the curriculum would be a denial of each teacher's right to follow her interests in [illegible]ing what she thought was best for each student in her classroom. Insofar as most teachers would see the basic skills of literacy and numeracy as essential for all their students, there would normally be no conflict between the right of every student to be taught the basic skills and a teacher's right to follow her inclinations. However, if student and teacher rights were in conflict on this issue, we would argue strongly that the right of the student must be paramount. As a consequence, the remainder of this paper will be restricted to consideration of those aspects of the curriculum which are specifically directed towards the attainment of literacy and numeracy by every student.

A school curriculum may, of course, be evaluated in a number of ways with many possible and in some cases conflicting criteria being used to determine the efficiency, effectiveness and the desirability of the curriculum being evaluated. One of the more important criteria in evaluating a curriculum is, surely, student outcomes, that is, what the student knows or can do as a result of being exposed to the curriculum. The measurement of specific student outcomes is perhaps the most appropriate criterion to use in evaluating a core curriculum such as the one described. It is in the areas of literacy and numeracy, in particular, that we wish to support movements towards measurement of specific student outcomes by means of criterion-referenced testing which are now taking place, to advocate the advantages of such tests for the purposes of curriculum evaluation and to outline a method for their use. The method outlined describes how the specific, essential student outcomes may be determined, how criterion-referenced tests to measure achievement of these outcomes may be developed, and how the results of such a testing program may be analysed.

## Criterion-Referenced Testing

Although numerous definitions and interpretations of criterion-referenced measurement and testing have been suggested in the past, most seem to consider that a criterion-referenced test attempts to determine the status of an examinee with respect to some well defined behavioural objectives (see for example, Millman, 1974). For this paper, we have adopted the definition first put forward by Glaser and Nitko (1971) who said:

> A criterion-referenced test is one that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specified performance standards.
>
> (Glaser and Nitko, 1971:653)

In addition, they go on to say that the performance standards are usually specified by defining some domain of tasks that the student should perform. Representative samples of tasks are taken and are used to make a statement about the performance of each individual relative to that domain. The specified performance standards, also known as criterion levels of a test, are often set arbitrarily at certain values which the examiner feels, in a sense, are minimum passing scores for the domain.

Thus the requirement in tests of basic skills is not to extend the range of scores as widely as possible to enable discriminations to be made between 'average' students nor, in this case, to enable the students at the upper end of the performance scale to show fully what they can do for the purposes of selection. The requirement is to discriminate between students at the lower end of the performance scale by establishing whether each student has achieved the specified performance level on a range of tasks considered essential for literacy and numeracy. The score distributions obtained by such tests are often highly negatively skewed which creates difficult problems in the analysis of results. These problems are discussed later.

Criterion-referenced tests may be constructed to assess performance on a limited number of tasks or they may be related to a wide spectrum of tasks. According to Millman (1973, 1974) the domain of tasks may form a single dimension or may be multidimensional, but in either case the structure of the domain must be precisely defined. If a test assesses performance on a number of dimensions within the behavioural domain, then the status of the student, should be determined separately for each dimension. In this case the multidimensional tests of, say, reading, writing and number work should be partitioned into sub-tests measuring more specific precisely-defined abilities which are necessary for literacy and numeracy, and performance on these sub-tests is examined.

In designing criterion-referenced tests in these areas it is first necessary to specify, as accurately as possible, the tasks and sub-tasks which a student must be able to perform if he is to be considered literate and numerate. Currently there are no objective criteria which can be used to select tasks and sub-tasks for inclusion in a list of tasks considered essential, consequently the inclusion or exclusion of tasks from such a list can only be a judgmental decision based on intuition or the best advice available. Before making decisions the test designer would be advised to consult all interested community groups because of the nature and importance of the student outcomes being measured. In particular, the views of teachers, students, parents, academics, employers and trade union officials should be sought and acted upon. Whether or not consensus is achieved, the test designer has a responsibility to make the decisions he has reached and the bases for them freely available for public scrutiny and criticism, so that his decisions can be evaluated subsequently both by further research and by further discussion within the community. It is probable that this evaluation would frequently have consequences for the curriculum designer as well as for the test designer. Ideally, of course, responsibility for both the curriculum and the testing would rest with the one individual, group or organization.

Once the essential tasks and sub-tasks have been identified, the development and selection of items to form sub-tests that measure performance on each sub-task is necessary. Moreover, a method is needed for deciding the level of competence on the sub-test required before performance of the sub-task is considered to be effective. The level can be made explicit by means of detailed objectives or it can be incorporated into the item design process implicitly rather than explicitly. The latter method, involving the use of items which define in operational terms the required level of competence, was used in the Australian Studies in School Performance and is described in greater detail in the first report of that study (Keeves and Bourke, 1976:29-30).

## The Concept of Mastery

A means of assessing whether each student has achieved an acceptable level of performance on each sub-task is required when the success or otherwise of a school's program is being evaluated. Clearly at one extreme, answering one half of the appropriate test items correctly would be inadequate, given the essential tasks identified. At the other extreme, to require the answering of all appropriate items correctly before performance was considered acceptable would be unrealistic. A performance level between these two extremes would seem to be required and the choice of the 80 per cent level, that is, where a student must be able to answer at least 80 per cent of all appropriate items correctly, is justified in Bourke and Keeves (1977:36). When a mastery learning program, such as those described in Block (1971), is being used, the agreed level of competence is required for a student to progress to the next unit of instruction as well as to indicate student performance for program evaluation purposes. However, where a learning program is not based on a mastery approach, it is still desirable to use the competence level for student and program assessment purposes. The proportion of students who reach the competence level set is a valuable indicator of a program's effectiveness.

Once the competence level is set for a particular sub-test, it is necessary to calculate the proportion of items which must be correct for a student to be considered to have achieved that level of competence. The items actually used to test a particular sub-task in a criterion-referenced testing program would ideally be a random sample of the hypothetical population of items which could have been used. However, the sample of items could also be chosen as representative, rather than a random sample, of all possible items. In the determination of a cutting score for each sub-test, it is not sufficient to use the competence level to simply calculate the corresponding proportion of items in the sub-test required to be correct for mastery, if one is intent on maximizing probabilities of making correct mastery/non-mastery decisions regarding student performance.

The purposes of identifying mastery and non-mastery students are normally to establish an index of success of a current educational program, and to obtain information for subsequent action either for individual students or for future curriculum changes. The proportions of students in the mastery and non-mastery categories will vary widely depending upon the cutting score chosen. Thus the calculation of a cutting score is a most important step requiring detailed consideration. Some of the key factors which influence the determination of mastery cutting scores are now identified and discussed.

It would seem possible and desirable to identify the relative costs in economic and in human terms of incorrectly withholding and incorrectly awarding mastery status. Several aspects would appear to be important. First, the cost of incorrectly awarding mastery status could be low because the student might subsequently learn how to do the task involved either incidentally or from another source, or the cost could be high because he might never learn to do the task. Secondly, when mastery status was incorrectly withheld, although the cost of providing unnecessary further instruction would be high, such instruction should be required only for a short period until it was realized that the student did not really require it. Thirdly, in addition to economic costs, psychological costs of mis-classification incurred by the student would be of great importance when the gains and losses of educational decisions were being considered. All identifiable gains and losses should be incorporated into the procedures used to determine a cutting score.

It is the obsolute number of correct responses which is crucial in a mastery test and guessing by students could assume importance, especially when multiple choice questions are used and the test is short. For a normative test the relative number of correct responses is the key, and consequently correction for guessing assumes lesser importance except in situations where the greater variance obtained is of benefit to the analysis, or where students have omitted responses to a substantial number of items. The effects of guessing on test scores have been well documented in the field of educational measurement, and clearly the expected number of correct responses to a test would differ according to whether or not an opportunity existed for guessing the correct answer. It would appear essential to incorporate an adjustment for guessing, where appropriate, in the calculation of a cutting score for a test assessing mastery.

In any testing situation it would be possible for a student to fail to record the correct response, not as a result of partial knowledge or of partially developed skill, but as a consequence of carelessness or what have been commonly identified as errors of measurement. Carelessness would thus be seen as a percentage of items incorrectly answered for which the correct response was known. It should be noted that for this type of responding behaviour the student could not be seen to be guessing, and situations where guessing was operating would differ from those where carelessness was operating. For constructed response items, even though guessing would largely be absent, carelessness might occur since it would be possible for the student to record a wrong response when he in fact knew the correct answer and would normally have made the appropriate response.

In some testing situations, it could occur that no prior knowledge existed about individual student performance on a test. In this case, the distribution of test scores for all students could be used as prior information, in addition to the individual student information available from the current testing. This is desirable in testing for mastery, because the score distributions are markedly skewed with a majority of students being above the mastery cutting score. Such information would assist in minimizing errors of classification made in assessing individuals as masters or non-masters. By taking into account such information when determining a cutting score, it is possible to assess the performance of individual students more meaningfully and to calculate the probabilities associated with that score more accurately.

It is important to examine the probability of making a correct or incorrect decision in classifying a student as a master not only at the defined cutting point but also at score levels above and below the cutting point. It would seem reasonable, when considering results for groups of students and when decisions regarding the advancement or retention of individual students were not being made that, at the cutting point, the probabilities of mis-classifying a mastery student as a non-mastery student, and vice versa, should be equated. At the cutting score, approximately half the students should best be considered to have mastered the task, because at this score the probabilities of being classified as master or non-master are approximately equal. At score values above and below the cutting score, the probabilities of incorrect classification would be lower than those at the cutting score.

Nevertheless, it must be acknowledged that for relatively short tests with few items the possibility of correctly classifying students might not be as great as with longer tests and there would be the risk that substantial proportions of students might be incorrectly classified on both sides of the cutting score. In assessing the performance of groups of students such errors might not be serious, but in cases where the individual performance of a student was being considered, errors of mis-classification might have more serious consequences and different criteria of classification might be advisable.

## The Calculation of Cutting Scores

A relatively complex procedure for calculating cutting scores which takes into account the factors just described is now outlined. Specific reference is made to a particular application of the method developed by Morgan for use in analysing results of Australian Studies in School Performance (Bourke and Keeves, 1977:279-296). The method employs a Bayesian, decision-theoretic approach to the determination of cutting scores which involves the calculation of the probability that a student's true score exceeds a set criterion score. The model incorporates collateral information about the distribution of test scores, has provision for correction for random guessing and carelessness, and assumes that the consequences of incorrectly classifying a student as a master would be equally serious as incorrectly classifying a student as a non-master. The value of using a complex statistical procedure is that errors of classification are minimised, the consequences of necessary but somewhat arbitrary judgments are exposed for scrutiny, and the probabilistic implications of classifying students with scores on either side of the cutting score are made clear. The mathematical development of the model is reported in Bourke and Keeves (1977:281-288).

The particular sub-tests used by Morgan to illustrate the usefulness of the model developed were the operations-on-numbers, reading simple sentences (picture items) and linguistic competence sub-tests. The tests are described in detail in Bourke and Keeves (1977:47-50 and 61-64) and the actual items used can be inspected in Bourke and Lewis (1976). The operations-on-numbers sub-test was part of the numeration tests and the other two sub-tests were sections of the reading tests given in 1975 to two groups of more than 6,000 students aged 10 and 14 years in normal schooling throughout Australia.

Table 1 Selected Informative Beta Prior Distributions for Tests.

| Origin of Prior | Age Level | Prior Distribution | Effective Prior Sample Size | Theoretical Distribution[c] | | | Empirical Distribution | |
|---|---|---|---|---|---|---|---|---|
| | | | | Mean | Mode | Standard Deviation | Mean | Mode |
| Reading[a] Test | 10 | B(8,2) | 10.0 | 0.800 | 0.875 | 0.121 | 0.786 | 0.862 |
| | 14 | B(11,2.5) | 13.5 | 0.815 | 0.870 | 0.102 | 0.823 | 0.889 |
| Numeration[a] Test | 10 | B(7,1.5) | 8.5 | 0.824 | 0.923 | 0.124 | 0.810 | 0.909 |
| | 14 | B(5.5,1) | 6.5 | 0.846 | 1.000 | 0.134 | 0.820 | 0.970 |
| Operations on Numbers[b] | 10 | B(2.5,1) | 3.5 | 0.714 | 1.000 | 0.213 | 0.762 | 1.000 |
| | 14 | B(8.5,1) | 9.5 | 0.895 | 1.000 | 0.095 | 0.918 | 1.000 |
| Picture Items[b] Sub-test | 10 | B(17,2) | 19.0 | 0.895 | 0.941 | 0.069 | 0.884 | 1.000 |
| | 14 | B(50,1) | 51.0 | 0.980 | 1.000 | 0.019 | 0.972 | 1.000 |
| Linguistic[b] Sub-test | 10 | B(9.5,1) | 10.5 | 0.905 | 1.000 | 0.087 | 0.918 | 1.000 |
| | 14 | B(24,1.5) | 25.5 | 0.941 | 0.979 | 0.046 | 0.973 | 1.000 |

[a] These refer to the total test.

[b] These refer to sub-tests.

[c] Values of statistics obtained using tables in Pearson (1968).

The distributions of scores on the total numeration and reading tests are used as prior knowledge of student performance on the three individual sub-tests considered. These observed scores have been translated into informative prior distributions which are summarized in Table 1. The table shows that each prior distribution has been characterized by a beta function with appropriately determined parameters. The parameters were estimated using the fractile assessment procedure (Novick and Jackson, 1974:167) which fits a function from the beta class to the observed values. Further, where necessary the estimated parameters were adjusted to conform with the statistics from the observed data. However it should be borne in mind that this procedure is at best a rough approximation, and that when observed distributions are highly skewed, the agreement between the estimated beta distribution and the observed distribution may not be perfect in any sense of the word. We may note that the use of a prior distribution with parameters a and b is equivalent mathematically to adding a + b items to the test (Novick and Jackson, 1974:126). Table 2 summarizes results of calculations of cutting scores based on the corrected model, for the three sub-tests. Three prior distributions were used at each age level and for each test. They were a uniform prior reflecting no knowledge of examinee's observed ability, a prior based on the scores of all students on the total reading and numeration tests, and a prior based on the observed performance of all subjects on the particular sub-test. Calculations were made for six criterion or mastery levels ranging from 0.70 to 0.95.

Table 2 Estimated Cutting Scores[a] for Tests Based on Three Different Prior Distributions and for Some Criterion Levels.

| Sub-test | Age Level | Length of Test | Posterior Distribution | Criterion Level | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
| Operations on Numbers | 10 | 8 | B(1+x,9-x)[b] | 6 | 6 | 7 | 7 | 8 | * |
| | | | B(7+x,9.5-x) | 4 | 5 | 6 | 6 | 7 | 8 |
| | | | B(2.5+x,9-x) | 5 | 6 | 6 | 7 | 8 | * |
| Operations on Numbers | 14 | 8 | B(1+x,9-x) | 6 | 6 | 7 | 7 | 8 | * |
| | | | B(5.5+x,9-x) | 4 | 5 | 6 | 6 | 7 | 8 |
| | | | B(8.5+x,9-x) | 3 | 4 | 5 | 6 | 7 | 8 |
| Picture Items | 10 | 5 | B(1+x,6-x) | 4 | 5 | 5 | 5 | * | * |
| | | | B(8+x,7-x) | 3 | 4 | 4 | 5 | 5 | * |
| | | | B(17+x,7-x) | 1 | 2 | 3 | 4 | 5 | * |
| Picture Items | 14 | 5 | B(1+x,6-x) | 4 | 5 | 5 | 5 | * | * |
| | | | B(11+x,7.5-x) | 2 | 4 | 4 | 5 | * | * |
| | | | B(50[c],6-x) | 0[d] | 0 | 0 | 0 | 0 | 3 |
| Linguistic Competence | 10 | 7 | B(1+x,8-x) | 6 | 6 | 7 | 7 | 7 | * |
| | | | B(8+x,9-x) | 5 | 6 | 6 | 7 | 7 | * |
| | | | B(9.5+x,8-x) | 4 | 5 | 5 | 6 | 6 | * |
| Linguistic Competence | 14 | 7 | B(1+x,8-x) | 6 | 6 | 7 | 7 | 7 | * |
| | | | B(11+x,9.5-x) | 5 | 6 | 6 | 7 | * | * |
| | | | B(24+x,9.5-x) | 1 | 2 | 3 | 4 | 5 | 7 |

[a] Estimated cutting scores reflect the situation when the probabilities of incorrectly awarding and withholding mastery status are approximately equal at 0.50 but in any case the corresponding probability is not less than 0.45.

[b] Beta distribution where x is the score on the test for a subject. For each test the first entry incorporates a uniform prior, the second the prior based on performance of all subjects on the total reading or numeration test, and the last the prior corresponding to performance of all subjects on the particular sub-test under investigation.

[c] To ease the effort of calculation the first parameter, a, for this posterior distribution was assumed equal to 50 for all values of x.

[d] A zero cutting score indicates that because of the assumed highly skewed prior distribution, B(50,1), which in effect swamps the likelihood, individual subjects are classified as masters on this sub-test irrespective of their score on the sub-test. In a sense, the test results for this situation have not contributed significantly for a better understanding of the subject's true performance level on this sub-test.

* Indicates that the probability of the observed ability being equal to or greater than the criterion level is less than 0.45 which was the lowest acceptable level.

Corresponding to each posterior distribution the cutting score is observed to remain constant or to increase in magnitude as the criterion level is increased. A higher criterion of performance demands a more stringent passing score on the test. In the examples presented in Table 2, the cutting scores derived on the assumption of an informative prior are generally lower than or equal to those derived when using a uniform prior. This is because the informative priors are based on all students' observed score distributions which are highly skewed, with modes near the top end of the scale. Moreover, prior distributions based on collateral information from the performance of all students on the sub-tests gave the consistently lowest cutting scores. Furthermore, it should be noted that the 14 year-old students were consistently better performers on the tests than the 10 year-old students.

In conclusion, we must stress that the meaningfulness of a cutting score is dependent on a variety of factors bound up with the test, including the view of expected performance levels held by the examiner for the students on the test. Given a test of a fixed length the examiner has to arrive at a criterion level very carefully. Moreover, as the results we have presented suggest, the examiner must be equally careful in choosing a prior distribution. Recent research has also suggested (Swaminathan, Hambleton and Algina, 1975) that the most appropriate prior to use is one composed of an individual prior for each examinee, together with a prior based on the collateral data specific to the test being used.

Clearly the model could be improved, though unfortunately it would as a result become more complicated. For example, the carelessness factor has been assumed to be constant for all students whereas it is almost certainly a function of proportion of correct responses achieved. The use of individual data for this factor as well as individual prior information rather than collateral information on test scores would lead to a more accurate determination of each student's mastery status, but would multiply the workload to an unacceptable degree, given the current state of the methods used to calculate cutting scores for criterion-referenced mastery tests.

Table 3 Percentages of Mastery Students for Sub-tests Based on Three Prior Distributions for Some Criterion Levels.

| Sub-test | Age Level | Test Length | Posterior Distribution | Criterion Level 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|
| Operations on Numbers | 10 | 8 | Uniform prior | 70 | 70 | 55 | 55 | 33 | * |
| | | | Collateral info. from total test | 87 | 80 | 70 | 70 | 55 | 33 |
| | | | Collateral info. from sub-test | 80 | 70 | 70 | 55 | 33 | * |
| Operations on Numbers | 14 | 8 | Uniform prior | 94 | 94 | 86 | 86 | 62 | * |
| | | | Collateral info. from total test | 98 | 97 | 94 | 94 | 86 | 62 |
| | | | Collateral info. from sub-test | 99 | 98 | 97 | 94 | 86 | 62 |
| Reading Simple Sentences | 10 | 5 | Uniform prior | 97 | 89 | 89 | 89 | * | * |
| | | | Collateral info. from total test | 99 | 97 | 97 | 89 | 89 | * |
| | | | Collateral info. from sub-test | 100 | 99 | 99 | 97 | 89 | * |
| Reading Simple Sentences | 14 | 5 | Uniform prior | 99 | 97 | 97 | 97 | * | * |
| | | | Collateral info. from total test | 100 | 99 | 99 | 97 | * | * |
| | | | Collateral info. from sub-test | * | * | * | * | * | 100 |
| Linguistic Competence | 10 | 7. | Uniform prior | 87 | 87 | 70 | 70 | 70 | * |
| | | | Collateral info. from total test | 93 | 87 | 87 | 70 | 70 | * |
| | | | Collateral info. from sub-test | 96 | 93 | 93 | 87 | 87 | * |
| Linguistic Competence | 14 | 7 | Uniform prior | 97 | 97 | 89 | 89 | 89 | * |
| | | | Collateral info. from total test | 99 | 97 | 97 | 89 | * | * |
| | | | Collateral info. from sub-test | 100 | 100 | 99 | 99 | 99 | 89 |

* Indicates that probability of classification errors are too high to be acceptable or that the cutting score required was zero.

The crucial nature of the decisions made as to the criterion level (that is, the level of competence) demanded and the posterior distribution used as prior information is amply illustrated in Table 3. In this table the percentages of students who would be considered to have mastered the sub-tests are shown for the criterion levels and posterior distributions used in Table 2 to show the cutting scores for the three sub-tests. The 0.80 criterion level and the collateral information on the total test score are the parameters that were actually used for these sub-tests and reported by Bourke and Keeves (1977). It can be seen that considerable differences in proportions of students classified as having achieved mastery would be obtained if other criterion levels or posterior distributions had been used. These wide fluctuations in the proportions of students achieving the competence level demanded are largely a function of the quite short sub-tests and the extremely skewed nature of some of the sub-test score distributions.

## Summary

It has been argued that there is a substantial component of the school curriculum which consists of essential knowledge and abilities which all students should possess if they are not to be disadvantaged in progress through the school system or in their everyday lives. The evaluation of this component, referred to as the core curriculum, should include measurement of student outcomes by means of criterion-referenced tests. An important objective within the core curriculum consists of the attainment of literacy and numeracy by every student.

Mastery learning and assessment would facilitate the achievement of at least minimum levels of competence, and the proportion of students achieving these levels would be an important indicator of the success of a school program. Cutting scores required for the assessment of mastery should minimise the probability of errors of classification of mastery and non-mastery students. Other important factors in determining cutting scores for each sub-test are the relative costs in economic and human terms of making incorrect decisions, errors due to such test factors as guessing and carelessness, and previous knowledge about test performance. A method of minimising the probabilities of incorrect decisions which incorporates consideration of each of the above factors has been outlined and an example of the use of the method has been described.

Finally, it is important to remember that the use of criterion-referenced tests and complex statistical procedures in establishing cutting scores does not eliminate the need for judgment at any stage in selecting the appropriate tasks for testing, in devising test items at an appropriate level of difficulty, or in setting competence levels. However, their use does identify and quantify the consequences of making judgmental decisions which are necessary in any evaluation exercise.

## References

Block, J.H. (Ed.) *Mastery Learning Theory and Practice*. New York: Holt, Rinehart and Winston. 1971.

Bourke, S.F. and Keeves, J.P. (Eds) *The Mastery of Literacy and Numeracy*. Canberra: AGPS. 1977 (in press).

Bourke, S.F. and Lewis, R. *Literacy and Numeracy in Australian Schools. Item Report*. Canberra: AGPS. 1976.

Glaser, R. and Nitko, A.J. 'Measurement in Learning and Instruction'. In R.J. Thorndike (Ed.) *Educational Measurement* (2nd ed.). Washington: American Council on Education, pp.625-670. 1971.

Keeves, J.P. and Bourke, S.F. *Literacy and Numeracy in Australian Schools. A First Report*. Canberra: AGPS. 1976.

Millman, J. 'Passing Scores and Test Lengths for Domain-Referenced Measures'. *Review of Educational Research*, 43:205-216. 1973.

Millman, J. 'Criterion-Referenced Measurement'. In W.J. Popham (Ed.) *Evaluation in Education : Current Applications* Berkeley, California: McCutchan Publishing Co. pp 311-397. 1974.

Novick, M.R. and Jackson, P.H. *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill. 1974.

Pearson, K. (Ed.) *Tables of the Incomplete Beta-Function* (2nd ed.). Cambridge: Cambridge University Press. 1968.

Smith, B.O. and Orlosky, D.E. *Socialization and Schooling : The Basics of Reform*. Bloomington, Indiana: Phi Delta Kappan. 1975.

Swaminathan, H., Hambleton, R.K. and Algina, J. 'A Bayesian Decision. Theoretic Procedure for Use with Criterion-Referenced Tests'. *Journal of Educational Measurement*, 12:87-98. 1975.

## FRIDAY 11 NOVEMBER

## 4.30 - 5.30pm

| **PAPERS** | **Rooms** |
|---|---|
| M. Martin, D. Symington (State College at Victoria, Toorak)<br>W. Harlen (Chelsea College, London)<br>Evaluation of School Based Curricula | Seminar Room B |
| J.R. Bainbridge (Blackburn, Vic.)<br>Class/Teacher Differences in Teacher, Curriculum, Institute Evaluation | Ladies Drawing Room |
| W.C. Hall (University of Adelaide)<br>An Approach to Evaluating Tertiary Courses | Seminar Room D |
| B.M. Bullivant (Monash University)<br>Curriculum Problematics in a Poly-ethnic Context | Meeting Room |
| J. Keeves, B. McBryde, I. Bennett (A.C.E.R.)<br>The Validity of Alternative Methods of Scaling School Assessment at the HSC Level for the Colleges and High Schools of the A.C.T. | Gateway Room |

# EVALUATION OF SCHOOL BASED CURRICULA

by

Marjory Martin, David Symington and Wynne Harlen*

State College of Victoria, Toorak
*Chelsea College, University of London

## Background to the model

To attempt to describe the factors which contribute to the educational climate in any school or school system is a task which obviously will lead to over-simplification. Yet it is essential that, while aware of the dangers inherent in the task, an effort is made to describe the context which has lead to the development of the model for evaluation of school based curricula to be presented in this paper. Four factors which are believed to be significant in creating the need for such a model will be discussed. Evidence can be cited to establish that these factors are presently operating in the Victorian Primary School System. However it is believed that they are not peculiar to Victoria but are operating to a substantial extent in many other school systems. It is not claimed that these are the only significant factors nor that they are independent of one another. Rather, they are chosen because they appear to be of greater significance than others and they are treated independently simply for ease of communication.

The first factor is the relationship between the individual school and the educational authority with respect to the school curriculum. In many school systems greater responsibility for curriculum decisions now rests with the school. For example for many years courses for State primary schools in Victoria were prepared by groups established by the central administration of the Education Department and these courses were mandatory for all State primary schools. During the past decade there has been a dramatic reversal of policy. The central administration has communicated to schools that it expects each individual school to accept responsibility for its own curriculum. The ultimate authority for the primary school curriculum in State schools continues to lie with the Minister of Education and is exercised through the Primary Schools Division of the Education Department and its inspectors. The shift in emphasis however is seen in the questions asked by these inspectors. Whereas in previous decades inspectors asked questions within the framework of the prescribed curriculum they now ask firstly that the school justify its own curriculum and then about the effectiveness of work within that curriculum.

The second factor to be noted here relates to the evaluation of the school programme. Part of the move to give schools greater autonomy and responsibility in curriculum matters includes the expectation that schools will play a more significant role in all aspects of evaluation. Again although this is believed to be a significant factor in many school systems the change will be illustrated by reference to the Victorian State school system. For most of the history of schools in Victoria evaluation of the school curriculum and inspection of the school by the district inspector could be equated. More recently however the view taken of evaluation of the school curriculum has been broadened. The Education Gazette and Teachers' Aid (1977) reflects this change in policy when introducing the concept of the school review. The topic is introduced by explaining who will participate.

> *'The process of evaluating and reporting upon a State primary school is known as a school review.*
>
> *As a general principle persons involved in the setting of school aims should be involved in the evaluation.'*
>
> *(page 155).*

The school review is presented as a joint exercise. The district inspector is still involved but now together with the staff of the school. Specific mention is made of the 'principal's or headmaster's written report and any information furnished by teachers.' (page 155) Evaluation of the school's programme is seen as an integral part of the activity of the school staff rather than the sole prerogative of the district inspector.

The third factor to be discussed is the changing role of the local community in the life of the school. In many places the general public is seeking and gaining a greater say in curriculum decision making. Legislative evidence of this factor can be found in the Education (Schools Council) Act 1975 passed by the Victorian Parliament. This Act has changed the function of the school committees in many areas but only one of these need concern us here. One of the additional responsibilities given to School Councils by the Act is to tender advice to the principal of the school on the general educational policy which might be followed in the conduct of the school.

Thus the individual school staff has been given greater responsibility in the area of curriculum planning and evaluation and at the same time been charged with communicating about the school curriculum with the local community to a greater degree through the school council. It is not surprising that many school principals and teachers are ill-prepared for some of these tasks that are being thrust upon them for the first time. This is particularly so in the area of evaluation. Teachers have for many years been responsible for evaluating the progress of pupils and reporting on this progress to parents either verbally or through school reports. They have not, however, been called upon to evaluate systematically their effectiveness at the level of individual classrooms or of the school as a whole. These tasks, as has been suggested earlier, were seen as the function of the district inspector. Yet, the school staff is now expected to evaluate systematically all aspects of the school curriculum, not only pupil progress but also the policy, programme and activities of the school. The July 1977 issue of C Scope, the newsletter of the Curriculum Standing Committee for Primary Education, includes a centrefold which sets out the current view of the committee on the curriculum. The section of the statement which deals with evaluation is included as Table 1 of this paper.

A fourth factor is part of a concern which originates independently of State legislature. The concern with 'standards' in education is felt not only by those outside schools. Those involved in education are looking more critically at their work and for ways of improving it. The increased range of new materials, methods and organizations open to schools also creates a need for information about what is being achieved in a school as a basis for decisions about changes which may be necessary or worth trying. There is a greater consciousness that decision making at all levels requires an examination of what is being done and that this should be an on-going process in a school if it is to respond to new developments and the challenge of preparing children to live in today's rapidly changing world.

## The start of the project

The project described in this paper arose from a notion that the methods which teachers would need for the task of evaluating systematically the school's curriculum already exist or can be devised by modifying existing methods. A Schools' Council publication, Harlen (1977), has gathered together, under the title Evaluation and the Teacher's Role, methods considered applicable for evaluation by teachers of the significant aspects of the school curriculum. In order to establish whether these methods could be applied and to explore the processes which facilitate their application, it was necessary in the first instance to work with practising teachers to consider and try these methods. Accordingly, by arrangement with a district inspector, six teachers from three schools met with the authors during July and August in a first attempt to develop a model for the evaluation of a school based curriculum. Two points should be made about the participating teachers. Firstly, the schools represented were those whose principals indicated interest in participating in the project. Secondly, by a fortunate co-incidence the teachers nominated by those schools covered a wide range of responsibility within the school, namely, a principal, vice principal, grade co-ordinator, subject co-ordinator and class room teacher.

Early in the group's discussions it was clear that a working definition of curriculum had to be agreed that went beyond the area of teaching materials and courses. The definition accepted as most suitable was 'everything that is deliberately organised by the school to impinge on the child and to effect the learning process.' Thus the group was concerned not only with the content of courses of study but also with the materials used by the teacher and the organization within the classroom and the school. In addition to this intended curriculum we were conscious of the 'hidden curriculum', that is the variety of experiences which are not planned, nor even necessarily acknowledged, but which may nevertheless result in the pupils learning attitudes and ideas as much as from the intended experiences.

The authors' interests in the project were to find out to what extent methods which already existed, or modifications of them, could be used by teachers and investigate how the process of school based curriculum evaluation might be implemented. By exploring with teachers the usefulness of various methods we hoped to be able to suggest a tentative model for school based curriculum evaluation.

Discussion with the teachers revealed two purposes they saw for the information gained by the use of the methods being proposed. Firstly as a basis for giving an account to parents, the school council, the district inspector and others of the school's goals and how it works towards them. The second use is as feedback to teachers themselves. The information could indicate areas where the teachers are implementing intentions successfully and others where intentions are not being fully implemented.

The starting point for thinking about the kinds of information needed for these purposes was to be clear about what are the intentions, or goals, of work in school. These goals were the criteria to be used in evaluation. It was found helpful to think in terms of goals of three closely related kinds: goals for the pupils, the teacher and goals for the school. The goals of the school as a whole include those for teachers and individual pupils, but extend beyond these in terms of things which concern the working of an institution and its relationship with the community around. Similarly goals for teachers go beyond those of pupils, extending to such things as co-operation with colleagues and interaction with parents.

At this point the group had to have some definite statements of goals to work with as a basis for evaluation. We were not attempting to prescribe goals or to dictate the form in which they should be stated. The particular goals of a school are a matter for that school and it was not the purpose of the project to go into the pros and cons of a common set of goals for all schools.

## Goals for pupils

The group started by thinking about goals for children. To find a set of goals which were acceptable to use as a basis for trying out evaluation methods, ideas were pooled and a list agreed as suitable for this purpose was drawn up. The list contained 37 statements, grouped under five headings. For example goals related to intellectual development included:

- *the child should have a wide vocabulary,*
- *the child should be able to use the four rules of arithmetic and basic mathematical concepts such as place value, number bases, units of measurement in solving problems,*
- *the child should be able to suggest explanations of observations and attempt to test his explanations,*

Goals related to aesthetic development included:

- *the child should know the appropriate techniques of some arts and crafts; for example, how to use paint, clay and to be able to use these to express his feelings or ideas,*
- *the child should be developing a personal appreciation of beauty in some of its forms both natural and artistic.*

Goals related to social/moral development included:

- *the child should know those moral values, relating to people and property, which are shared by the majority of members of the society,*
- *the child should be developing tolerance; respecting and appreciating others, their feelings, views and capabilities.*

**Goals related to physical development included:**

- *the child should know how to apply the basic principles of health, hygiene and safety and should know what to do in emergencies,*
- *the child should be able to control his body in a range of movements (including swimming) both as an individual and in organised games and sports.*

Goals related to emotional/personal development included:

- *the child should be developing ability to plan and perform his work independently and responsibly, attempting to overcome problems himself before asking for help,*
- *the child should be developing the capacity to form a considered opinion and to act upon it even if this means rejecting conventional thought and behaviour.*

These pupil goals were used as the criteria against which the adequacy of the experiences and opportunities provided for the children could be judged. For example the goal 'the child should have some elementary knowledge of the characteristics, needs and variety of living things' has definite implications for some of the content and nature of a child's experience at school.

## Evaluation at the level of the class

The next step to be taken after clarifying the goals was to ask the question "What is done in school to give children a chance to achieve each goal?" This was put *before* the question "Are the children achieving the goals?" The reason was not because it is not considered important to evaluate children's achievement, but because it should first be clear as to whether or not opportunities are there before looking for any affect. If the children never encounter living things for example, how can one expect them to develop knowledge about them? In such a case, to assess the children's knowledge of living things would not tell anything about what they had learned in school.

We therefore looked first at what could be done to evaluate, at the level of the class, the opportunities provided for children in relation to each goal. Since content of activities is not the only aspect which can influence children's learning, we looked at the opportunities under these headings:

Content
Teaching methods and organisation
Physical organisation of the class

We then attempted to describe what was done for each goal, under one, two or all three of these headings where appropriate. For instance consider again the goal:

> *The child should have some elementary knowledge of the characteristics, needs and variety of living things.*

The possible entries under these headings might be:

Content:

The life cycle of plants grown from seeds or bulbs in the classroom. The seasonal variations in plants around the school. The feeding habits and sensitivity of small animals and minibeasts kept in the classroom. Knowledge of other animals and plants seen on visits, TV or in pictures.

Methods and Organisation:

Investigation of plants and animals in classroom, planned and carried out by children. Nature charts, nature table, walks, visits, TV.

Physical Organisation:

Containers readily available for plants and animals brought by children. Display space for charts. Nature table where children can observe. Bulk loan of books from library accessible for reference.

Again it must be emphasised that we were describing *what was done* not what we thought should be done. It was necessary to be honest and frank with oneself; in that way we became aware of what was not done that ought to be done, and sometimes vice versa. For instance, an honest description of what is done for this goal in another class might be:

Content:

Incidental - what the children bring in or what is on TV.

Methods and Organisation:

Looking at items brought in. Set work to be done in library after TV programmes.

Physical Organisation:

Pots available for flowers and bulbs.

Just describing what is done is not evaluation, however, even though it may be a revealing exercise in its own right. For evaluation it was necessary to ask two kinds of questions about the description of what was done:

Firstly, is what is described appropriate and adequate for achieving the goal it was supposed to achieve? *(the descriptive evaluation)*

Secondly, how is what is observed in practice related to what was described and is what happens appropriate and adequate for achieving the goal? *(the observational evaluation)*

For the *descriptive evaluation* we found it necessary to ask certain questions about the provision described for each goal. The questions were not intended to provide a yes/no checklist, but rather to suggest aspects which should be considered in judging the adequacy of what was done.

As an example, for the goal about knowledge of living things we proposed asking:

What things are displayed alive in the classroom at school?

What changes are made in these displays?

What opportunities do children have outside the school to see a variety of living things?

To what extent are children involved in looking after animals and plants as a regular activity?

Is it possible for children to watch the whole life cycle of animals and plants?

What books, charts, films, etc. are available for children to broaden their knowledge of living things?

How are the experiences provided in each grade co-ordinated with those in other grades to allow continuity in development of knowledge and sensitivity towards living things?

Such questions we thought could be the basis for a teacher to ask herself, or himself, about the adequacy of what is done to achieve this goal. They might also be the kinds of questions an informed outsider might ask.

The second point of evaluation we call *observational evaluation* because it is concerned with judgements made in practice rather than in theory. A teacher may be convinced that she gives children, for instance, opportunity to talk about their ideas on a certain subject, but someone observing the class may see the teacher giving *her* ideas only.

For this evaluation a teacher needs the help of someone willing to observe the class in action and to give their account of what is seen. For teachers used to teaching with others present and discussing their work with them this is nothing new and the only extra work involved would be agreeing the framework for observation and the criteria to be used. For those not used to this situation there need be no stress if they are serious in their desire to find out things which will help them improve their teaching. The purpose is not for the observer to 'examine' or 'inspect' but for a teacher to have access to another point of view about what is happening in the classroom - something that he or she cannot do alone. The information the observer can provide will be found by looking at, for example:

- what the teacher does as well as listening to what she says;
- the children's reactions and discussing the work with a few individuals;
- the physical organisation - how this inhibits or facilitates what the teacher is trying to do;
- the opportunities provided for pupils by the materials and the methods used.

To sum up, the ways we suggested for answering the question about opportunities given to children for achieving the goals involve a teacher in:

1. describing what is done,
2. probing the adequacy and appropriateness of this, by asking herself certain questions,
3. inviting an observer's view of it.

Putting the product of these processes together would, we think, increase awareness of what a teacher is doing to help her children achieve the goals she has for them, and enable her to identify where this may not be as satisfactory as she would like.

In line with our purpose of exploring methods rather than developing whole evaluation instruments, we worked through a sample of goals until we were satisfied that a useful approach to evaluation at this level had been found. Through the experience of doing this we were not only able to adapt methods so as to be practicable but we arrived at another, perhaps more important, finding. This was that the process of attempting to develop methods and particularly that of identifying evaluation questions was found by the teachers to be a most rewarding task. We became convinced that had we presented to the teachers a ready-made set of evaluation questions, such as the ones we worked out, there would have been much less value for the teachers. It is difficult to substantiate this claim except to report that in reflecting on the work the teachers volunteered that the process of thinking out the questions which they should be asking themselves gave them an understanding of the whole evaluation process. Without this the product if presented to them, might well have been regarded as just another form to fill, or questionnaire to answer, whose significance for self-evaluation would have been minimal."

## Evaluating pupil progress

Having looked at ways teachers can evaluate the opportunities they are providing for pupils to achieve the set goals we turned our attention to the question "Are children achieving the goals?" The goals we had used as examples were of a general kind, applicable to the whole primary stage of education. It was first necessary to break each of them down to identify a sequence of steps relating to behaviour (actions and responses of all kinds) which could be readily observed. For some of the goals these descriptions could then serve as a framework for evaluating children's progress by observing them during their normal work.

The breakdown of the goals into identifiable steps can be illustrated by showing the four steps we devised in relation to the goal "knowledge of characteristics, needs and variety of living things."

(i) Recognises only a few familiar animals as being alive, but does not see plants as living.

(ii) Correctly identifies common plants and animals as living and recognises the life cycle of many of them.

(iii) Knows that animals and plants exist in wide variety, all having some way of feeding, developing and reproducing themselves.

(iv) Knows the characteristics of living things and that all require energy (food) which they obtain and use in a variety of ways.

The number of steps we identified varied according to the goal. For instance only three steps were found useful for describing development towards the goal, "The child should be able to discuss with tolerance opinions and values of others which are different from his own." For the goal "The child should be able to use the four rules of arithmetic and basic mathematical concepts such as place value, number bases, units of measurement in solving problems" it was found necessary to divide this into two aspects, pure and applied, and six steps within each were identified.

The developmental sequences defined in this way could be used to assess and record children's attainment during their progress through school. The description of these developmental sequences provides a method of recording development in relation to *all* the goals, rather than, as happens at present, only in relation to skills and abilities which are easily tested. The steps provide criteria for assessment so that judgements made by one teacher can be related to the same criteria as judgements made by another. This means, of course, that the criteria must be agreed by all those concerned - a major reason for suggesting that a school staff work out their own goals and criteria for development towards them. However, in doing this it is useful to have access to the ideas of others and we have worked out descriptions of development towards most of our 37 goals. It was noted that information could be gathered by tests if the levels of attainment were not evident in the pupils' normal work.

## Evaluation of the school as a whole

The third level at which we looked at school-based evaluation was the level of the school as a whole. As stated earlier, the school as an institution has goals which include those of the work of teachers and the goals for the pupils. So what we have discussed to this point is part of the whole school evaluation but now we looked at other aspects not covered by evaluation at the class and pupil levels.

For evaluation of the school, information is needed about all aspects. But what are these? We identified ten facets of the school as a whole which could be used as headings for gathering different kinds of information.

1. Basic statistics relating to pupils and staff.
2. Work programme policy.
3. Record keeping policy.
4. Class organisation policy.
5. Procedures for decision-making; staff duties and responsibilities.
6. Staff meetings and staff development policy.
7. Provision of supplementary programmes: remedial, enrichment, optional.
8. Attainment and diagnostic information.

9. Physical resources for teaching
   (a) indoor, (b) outdoor.

10. Programme for parental and community involvement and communication with the School Council.

Under each of these headings there are a number of questions that can be asked so that facts are gathered to answer them. For example, the questions we suggested in relation to "Procedures for decision-making; staff duties and responsibilities" are:

How are areas of staff responsibility defined?

What is the mechanism for deciding these areas of responsibility and modifying them when staff leave, or at other times?

To what extent are the preferences of staff members taken into account?

How are staff made conversant with the structure of responsibilities?

How can individual teachers have opportunity to assume responsibilities of a higher order?

Do responsibilities relate to the individual strengths of teachers or to seniority?

When were responsibilities last reviewed?

What are the duties of teachers with special responsibility for subject areas? for year group co-ordination?

etc.................

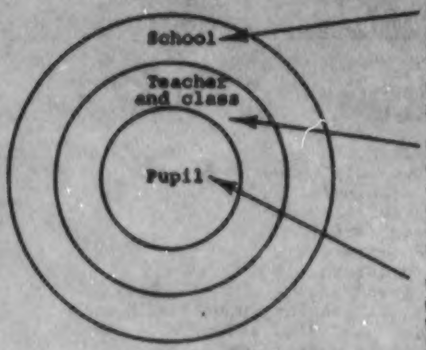
But stating what is done is only one step towards evaluation. Those within the school may not feel that what is done is sufficient or suitable. It is necessary to gather ideas and opinions about what changes might be made and discuss action which should be taken. This can be done by circulating to individual members of staff a form such as the one illustrated.

| Question about aspects of the school | Statement of what is done | Suggestions of changes which should be made |
|---|---|---|
| | | |

with the first two columns completed, asking for comments in the third column. Alternatively a special meeting, or conference, of staff could tackle the evaluation together.

Not all aspects need to be considered in this way at any one time - this would probably be too great a job - but rather the sections relevant to current concerns used as required. Thus record keeping could be reviewed in one term, parental and community involvement in another, and so on. However, the information about what is done should be kept up to date at all times, and this may suggest which aspects are due for detailed review.

Finally we can summarise the approach to School Based Curriculum Evaluation which we have adopted as attempting to work out ways of gathering the information needed to answer the following questions:



What are its goals, policies and procedures? What is done at present in relation to these? What changes should be made?

What is done through content, methods and organisation to help achieve pupils' goals? Is it adequate and appropriate?

What are the goals of their work? What point in development towards goals have they reached?

## Conclusion

The purpose of the project was to examine and trial with a group of teachers existing methods of gathering information or modifications of them and to determine what was acceptable to teachers for the evaluation of school based curriculum.

As a result we saw the kinds of processes and approaches at which we arrived, and described here, as a possible model for use by teachers. In particular we found that teachers' involvement in working out questions to ask in their own school is an important part of the process of school based evaluation. This was seen to be not only a means of developing instruments focussed on the particular problems of an individual school, but also the route to involving teachers in their own evaluation.

Much remains to be done in testing the model developed and it is our intention to assist schools interested in working through the model. We have not developed a package for use by schools but we have assembled a set of methods appropriate to the evaluation of school based curricula. The value of the model for schools will come through teachers working with these methods.

## References

Curriculum Standing Committee for Primary Education. C. Scope, Number 2, July 1977.

Education Department of Victoria (1977). Primary school reviews, *Education Gazette and Teachers' Aid*, 77, 2, 155-156.

Harlen, W. (Ed.) (1977) *Evaluation and the Teacher's Role.* To be published in the Schools' Council Research Series by Macmillan, London.

## TABLE I

## EVALUATION

| | |
|---|---|
| EVALUATION OF POLICY | -Appropriate<br>-Relevant<br>-Clear<br>-Consistent |
| EVALUATION OF PROCESS | -Involves all relevant people in a meaningful way<br>-Effective methods:<br>-cohesive<br>-comprehensive<br>-practical<br>-involving<br>-Questionnaires and checklists<br>-Meetings and discussions<br>-Face to face techniques e.g. Interviews |
| EVALUATION OF ORGANIZATION | -Curriculum Program<br>Comprehensive, consistent, realistic, theoretically sound.<br>-Staff<br>Working as a team, clear about goals, communicating adequately, utilising skills and abilities.<br>-Children<br>Maximise use of resources, Ensure optimum learning conditions.<br>-Resources<br>Available, accessible, used fully and effectively. |
| THROUGH | -As above<br>-Unobtrusive measures<br>-Classroom observation. |
| EVALUATION OF | -Learning outcomes<br>-Classroom interactions<br>-Teacher effectiveness<br>-Student progress (In the light of policy and program) |
| THROUGH | -A wide range of evaluative techniques including:<br>Observational techniques<br>Direct measures e.g. 'cloze' procedure<br>Teacher's ratings<br>Criterion-Referenced Tests<br>Student self-reports, etc. |

# CLASS /TEACHER DIFFERENCES in TEACHER, CURRICULUM, INSTITUTE EVALUATION.

Paper prepared for the -
AUSTRALIAN ASSOCIATION for RESEARCH in EDUCATION
ANNUAL CONFERENCE, CANBERRA. 11 - 13th. 1977.

J. E. BAINBRIDGE.

## I PROLOGUE.

In my variable Consulting arrangements with Monash University Computer Centre, I discuss all kinds of problems in the interpretation of all kinds of data on all kinds of subjects. These always involve the nature of the data, its collection, and the Investigators' hypotheses as to the relationships in the technical field being studied. To me this is the essence of Applied Statistics. It also means that no two cases are identical!

In this paper I discuss two cases from my case book in order to illustrate these basic facts, and some methods of dealing with them. The examples seem relevant to the Conference Theme of Curriculum Evaluation, and also to a discussion at a recent HERDSA conference on "Measuring the Performance of Institutions" - a touchy and difficult subject which was not pursued to its logical conclusion. There, and elsewhere, there seems to be a reluctance to tackle the problems of "Measurements which will be Used" I attempt a contribution to that topic.

## II CASE 1.

This case concerns a rather large secondary school which was well respected, had 8 more-or-less parallel classes in its upper years, and which had adequate access to a computer. It wished to teach computing to each of these 8 classes. The vital question was -

Could all these classes be taught computing?

There were complications. There were approximately 200 students involved. They were allocated to classes according to the courses they were expected to follow for the H.S.C. Examination. This meant that one class was made up of those specialising in Maths., another of those taking Languages, and so on. This division had its roots in matters of time-tabling. The classes differed in their interests, and in their knowledge in subjects such as Mathematics or Languages. The school had measures of the abilities of each scholar, and these, besides emphasising the differences between classes, showed that the eight classes fell into two distinct ability groups which will be referred to as the higher, and lower ability groups - see Table I.

TABLE I

Classes, Abilities, and Teachers.

| Lower Ability Classes | | | Higher Ability Classes | | |
|---|---|---|---|---|---|
| Class | Av. Ability | Teacher | Class | Av. Ability | Teacher |
| I | 40 | *A | V | 73 | D |
| II | 45 | B | VI | 79 | C |
| III | 45 | C | VII | 83 | E |
| IV | 44 | B | VIII | 65 | *F |

* See discussion in text.

Six Teachers, A - F were available for teaching, and were allocated as shown in Table I. Teacher F was an experienced maths. teacher, and would teach those specialising in mathematics at H.S.C. He was to teach computing to these same students. Teacher A argued strongly that class I, with low mathematical ability, should be taught computing in a non-standard way which he outlined. The powers that be said "Go to it and see how you get on!" Other teachers followed the standard syllabus and curriculum. There being only six available and qualified teachers, two had to take two classes each, and timetabling settled their allocation.

The School was not interested in comparisons between the upper and lower ability groups, but was interested in teacher and curriculum effects. These facts determined the method of analysis of the data.

Two outcomes were measured for each student - (a) His ability in Computing, and (b) His attitude to Computing at the end of the course as compared with his attitude to mathematics at the start.

## II (a) Analysis.

It was expected that success at computing would be related to ability - as assessed by the school. Because average abilities differed widely from class to class, it seemed futile to compare class achievements, either for indicating curriculum success, or for teacher success, unless correction could be made for the effect of ability differences This effect could be estimated from the relation between computing success and the school assessment of ability amongst students in the same class. This was possible.

## II (b) Synthetic Data.

Some hypothetical data was constructed for 4 classes, with 4 students per class, and with simple numbers, so that readers may, if they will, easily verify the calculations and inferences. The data contain variations in school assessed ability, in class (and in syllabus in the case of class I), and in the computing achievement of each of the hypothetical students. The data are given in Fig. 1.

> Note! The data are not real. Real data will be discussed later. It is often helpful to use synthetic data, and see if one's analytic method recovers the known structure of the data.

The symbol y is used to represent the success at computing, i.e. the outcome of the exercise. The symbol x is used to represent an input affecting learning, e.g. x1 is the schools assessment of a scholars ability, and x2 is coded to represent his class, i.e. the learning environment which he enjoyed. The reader will notice that Fig.1 contains values not of x, and y, but of $(x-\bar{x})$ and of $(y-\bar{y})$ where $\bar{x}$ and $\bar{y}$ are the averages, or means, of the x's and y's This follows usual practice, and is discussed by Yule(1). It simplifies arithmetic. It also gives expression to the essential humility of Statisticians, who wish to study data without reference to their preconceived opinions - e.g. that zero marks is bad, and 100 is perfect. All individual scores are related to the mean for all students.

The basic hypothesis is that success in computing $(y-\bar{y})$ increases as the school assessment of ability $(x1-\bar{x}1)$ increases, and also that students in class I will do better than students in classes II, III, or IV who have the same innate ability. These can be expressed by the relation:

$$y-\bar{y} = b1(x1-\bar{x}1) + b2(x2-\bar{x}2) \qquad \text{Eq. (1)}$$

Yule(1) set out in his "Introduction to the Theory of Statistics" of 1911, to provide a systematic course on statistical methods suited to the needs of those who possess only a limited knowledge of mathematics. In this he was so successful that the book ran into 10 editions. In his chapter on simple curve fitting, he made the point that a simple equation such as (1), cannot exactly agree with any large amount of any imprecise data, and that the coefficients b1, and b2 must be adjusted to give a best fit. He discusses the meaning of 'best fit', and how the values of b1 and b2 should be calculated. He showed that the equations are:

$$\begin{aligned} b_1\Sigma(x_1-\bar{x}_1)^2 + b_2\Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2) &= \Sigma(x_1-\bar{x}_1)(y-\bar{y}) \\ b_1\Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2) + b_2\Sigma(x_2-\bar{x}_2)^2 &= \Sigma(x_2-\bar{x}_2)(y-\bar{y}) \qquad \text{Equ. 2} \end{aligned}$$

In 1937 Yule asked a younger man, M.G.Kendall, to assist him in updating his book. The revision was published in 1938 with the same title, but with joint authorship, Yule and Kendall, and was followed by two further minor revisions numbers 12, and 13. The 14 th. edition, which has had at least 3 impressions, was published in 1958, just after Yule's death at the age of 81.

The important topics added in these last 4 editions were the concepts of the Analysis of Variance (he used a different name), a very clear and exact account of Significance, and of its application via t and F tables. The solution of the simultaneous equations involved in calculating $b_1$ and $b_2$ follows an earlier treatment by R.A.Fisher(2) which Yule described as straightforward, but requiring rather strenuous arithmetic! Some useful extensions to Fisher's work were published in 1938 by W.G.Cochran(3) who recognised that the coefficients of the left-hand side of his calculations formed what mathematicians had long called the 'inverse' of a matrix. These same methods were followed by O.L.Davies(4) who edited "Statistical Methods in Research and Production" on behalf of Imperial Chemical Industries Ltd, and a number of members of their Staff. W.G.Cochran and Gertrude Cox(5) also used the same plan.

By the late 1950's computers had arrived. One no longer needed arithmetic - the computer would do it! One need understand no Statistics either, a computer program would perform any standard computation. Applied Statistics took off at high speed, with no helmsman and promptly ran aground!

The first hazard is illustrated, in elementary form, by the data plotted in Fig. 1. The primary values to be calculated are -

$$\begin{array}{lll} \Sigma(x_1-\bar{x}_1)^2 & \Sigma(x_1-\bar{x}_1)(x_2-\bar{x}_2) & \Sigma(x_1-\bar{x}_1)(y-\bar{y}) \\ & \Sigma(x_2-\bar{x}_2)^2 & \Sigma(x_2-\bar{x}_2)(y-\bar{y}) \\ & & \Sigma(y-\bar{y})^2 \end{array}$$

whose values are -

$$\begin{array}{ccc} 590 & -96 & 398 \\ & 48 & 0 \\ & & 470 \end{array}$$

To find the regression of Computing Success on Pre-ability only, ignore $x_2$ entirely and calculate $b_1=\Sigma(x_1-\bar{x}_1)(y-\bar{y})/\Sigma(x_1-\bar{x}_1)^2 = 398/590 = 0.675$

The Regression Sum-of-Squares is $b_1\Sigma(x_1-\bar{x}_1)(y-\bar{y}) = .675 \times 398 = 269$

The Residual Sum of-Squares is $470 - 269 =$ 201

If we use both predictors, we substitute values in Equations 2 as follows -

$$590.b_1 -96.b_2 = 398$$

and

$$-96.b_1 +48.b_2 = 0$$

doubling the 2nd. equation, and adding, gives -

$$398.b_1 = 398 \quad \text{or} \quad b_1=1.0$$

$$\text{and} \quad b_2=2.0$$

The Regression Sum-of-Squares is $b_1\Sigma(x_1-\bar{x}_1)(y-\bar{y})+b_2\Sigma(x_2-\bar{x}_2)(y-\bar{y})$

or $1.0 \times 398 + 2.0 \times 0.0 = 398$

The Residual Sum-of-Squares is $470 - 398 = 72$

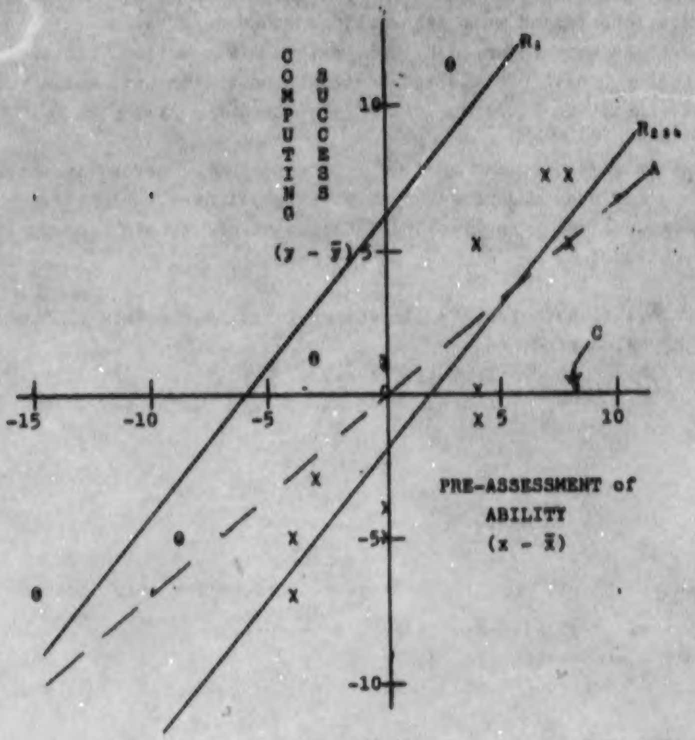
The Residual Degrees of Freedom $= 13$

The Estimate of the Residual Variance, or Mean Square $= 72/13 = 5.55$

Substitution in Equation 1 shows that the Computing success of any scholar, measured from the mean of all scholars, is given by 1) his assessed ability above the mean for all scholars, plus 6 if he is in class 1, or less 2 if in another class. So much for the synthetic data and its analysis.

It is sometimes stated that Multiple Regression Analysis is not applicable to "quantal" or non-quantative data. Here the Class to which a student belongs is certainly quantal, but the analysis seems trouble-free. This is partly due to the choice of values of x used, i.e. $x_2=3$ for class 1, and $x_2=-1$ for all other classes. Because there are 3 times as many scholars in "all other classes" as in class 1, $\Sigma x_2=0$, $=\bar{x}_2$, predicted values are the normal values for the scholar's ability plus $b_2(x_2-\bar{x}_2)$, or $b_2x_2$. For class 1 this is (normal + $3b_2$) and for other classes (normal - $b_2$). The hypothesis about classes is NOT that ahey form a linear sequence, and they should NOT be scored linearly - e.g. 0,1,2,3.

FIG. 1



SYNTHETIC DATA.

Deviations from Means

| | | |
|---|---|---|
| Outcome | (Ordinate) | Computing Success |
| Predictor | (Abscissa) | Pre-Assessment of Ability |
| | (Class) | Class 1 denoted by θ |
| | | Classes 2,3&4 denoted by X |

| | | | | | |
|---|---|---|---|---|---|
| Line | C | Regression of | Success | on | Class Average. |
| Line | A | | | | Pre-Ability |
| Lines | R | | | | Both Predictors. |

## TABLE II LOWER ABILITY GROUP.

Copy of Computer Output - Excluding Dispersion Matrix, Variables 7, 8, and 9, and some variable names.

```
Predictor Variables.                                        Averages
  Variable (1) is   Pre - Attitude to Maths.                  39.51
           (2) is   Pre - Achievement in Maths.               43.22
  Variables (3), (4), & (5) take the following values according
           to the class and teacher involved.
      Class      I       11      III      IV
           (3) 3.0    -1.0    -1.0    -1.0                    0.00
           (4) 0.0    -1.0     0.0     1.0                   -0.05
           (5) 0.0     1.0    -2.0     1.0                   -0.03

           (6)      Interaction of (1) & (2)                 602.33

  Outcome Variables
    Variable (10)   Post- less Pre-Attitude to Mathematics.    4.52
             (11)   Post-Achievement in Computing.             5.43


Number of Entities read = 96.
Analysis of Variance, and Regression Analysis.

                                                                      Variance Sign-
Predictor  Outcome          Regression Sum/Squrs.        R e s i d u a l     Ratio  ific-
                            Previous  This Factor     S/S    D/F Mean Sqr.      F   ance.
    I        10                0.000     4991.898  27988.060  94 297.74532    16.77  ***
             11               0.0000      42.6720  1322.8176  94 14.072527     3.03

    2        10             4991.898     3201.219  24786.841  93 266.52517    12.01  **
             11              42.6720       1.4295  1321.3881  93 14.208474     0.10
Regression:  Predictor   1         2                                    1         2
                 b   -0.33106 -0.51646             Outcome    b     0.040618  0.010914
    Outcome     S/S    2163.1   3201.2               11      S/S      32.562    1.4295
      10        Sig.     **       **                         Sig.

    3        10             8193.118     3050.164  21736.677  92 236.26822    12.91  ***
             11             44.10150    476.11799  845.27009  92 9.1877184    51.82  ***
    4        10            11243.282      202.564  21534.113  91 236.63860     0.86
             11            520.21949     16.43450  828.83560  91 9.1080835     1.80
    5        10            11445.846     1190.286  20343.826  90 226.04252     5.27  *
             11            536.65399      0.01319  828.82241  90 9.2091378     0.00

Regression:  Predictor     1         2          3         4        5
                 b     -0.44149  -0.38644     3.4206    1.7973  -2.8534
    Outcome     S/S      3497.2    1670.0     3102.1    143.77   1190.3
      10        Sig.      ***        **        ***                 *

                 b     0.010715  0.056726     1.3378  -0.60533  0.009499
    Outcome     S/S      2.0599    35.985     474.50    16.307  0.013191
      11        Sig.                  *        ***
```

## TABLE III.

### Higher Ability Classes.

| Predictor Variables | | Averages |
|---|---|---|
| Variable (1) is | Pre - Attitude to Maths. | 55.20 |
| (2) is | Pre - Achievement in Maths. | 76.36 |

Variables (3), (4) & (5) take the following values according to the class and teacher involved.

| Class | V | VI | VII | VIII | |
|---|---|---|---|---|---|
| (3) | -1.0 | -1.0 | -1.0 | 3.0 | -0.06087 |
| (4) | -1.0 | 0.0 | 1.0 | 0.0 | 0.017391 |
| (5) | 1.0 | -2.0 | 1.0 | 0.0 | 0.034783 |
| (6) | Interaction of (1) & (2) | | | | 58.165 |

| Outcome Variables | | |
|---|---|---|
| Variable (10) | Post- less Pre-Attitude to Maths. | -0.817 |
| (11) | Post-Achievement in Computing | 11.930 |

Number of Entities Read = 115

Analysis of Variance, and Regression Analysis.

| Predictor | Outcome | Regression Sum/Squrs. Previous | Regression Sum/Squrs. This Factor | Residual S/S | Residual D/F | Residual Mean Squ. | Variance Ratio F | Significance. |
|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 0.000 | 1558.639 | 12368.526 | 113 | 109.45598 | 14.24 | *** |
| | 11 | 0.0000 | 63.9312 | 1945.5123 | 113 | 17.216923 | 3.71 | |
| 2 | 10 | 1558.639 | 208.176 | 12160.350 | 112 | 108.57455 | 1.92 | |
| | 11 | 63.9312 | 42.0222 | 1903.4901 | 112 | 16.995447 | 2.47 | |

| Regression: | Predictor | 1 | 2 |
|---|---|---|---|
| Outcome 10 | b | -0.24368 | -0.13639 |
| | S/S | 868.67 | 208.18 |
| | Sig. | ** | |

| | Predictor | 1 | 2 |
|---|---|---|---|
| Outcome 11 | b | 0.082706 | -0.06128 |
| | S/S | 100.07 | 42.022 |
| | Sig. | * | |

| Predictor | Outcome | Regression Sum/Squrs. Previous | Regression Sum/Squrs. This Factor | Residual S/S | Residual D/F | Residual Mean Squ. | Variance Ratio F | Significance. |
|---|---|---|---|---|---|---|---|---|
| 3 | 10 | 1766.815 | 8.795 | 12151.555 | 111 | 109.47347 | 0.08 | |
| | 11 | 105.9534 | 145.3524 | 1758.1377 | 111 | 15.839078 | 9.18 | ** |
| 4 | 10 | 1775.611 | 463.851 | 11687.704 | 110 | 106.25186 | 4.37 | * |
| | 11 | 251.3058 | 65.1302 | 1693.0075 | 110 | 15.390977 | 4.23 | * |
| 5 | 10 | 2239.461 | 1039.939 | 10647.765 | 109 | 97.68592 | 10.65 | ** |
| | 11 | 316.4360 | 10.6714 | 1682.3361 | 109 | 15.434276 | 0.69 | |

| Regression: | Predictor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Outcome 10 | b | -0.25895 | -0.13514 | 0.10890 | 2.8569 | -2.4713 |
| | S/S | 958.1 | 130.95 | 2.5488 | 467.00 | 1039.9 |
| | Sig. | ** | | | * | ** |
| Outcome 11 | b | 0.069317 | 0.002008 | 0.76091 | 1.0679 | -0.25034 |
| | S/S | 68.655 | 0.028914 | 124.43 | 65.249 | 10.671 |
| | Sig. | * | | ** | * | |

A second point of note is the order of analysis. Had the first regression been on class only, one would initially ignore $x_1$ and write -

$b_2 = \Sigma(x_2-\bar{x}_2)(y-\bar{y})/\Sigma(x_2-\bar{x}_2)^2 = 0/48 = 0$

The Regression Sum-of-Squares $= b_2\Sigma(x_2-\bar{x}_2)(y-\bar{y}) = 0$ and we have made zero progress. Note the two conditions - $x_2$ considered alone and first, extra regression sum-of-squares Nil,
$x_2$ following $x_1$, extra regression sum-of-squares $=398-269 = 129$.

It is usually true that the improvement in fit of the regression equation to the data, for any one predictor, depends upon which other predictors have preceeded it. This makes order important both for Prediction, and also for Understanding. One may suggest that the crucial effect, in which one is interested, should follow all those other effects which might interfere with its interpretation. This requires two qualifications which will be discussed later in connection with the real data.

## II (c) CASE I: Real Data.

The differences between Real and Synthetic data are -

1) The Data are Real and accordingly statistical conclusions are conclusions about Education in the real world, and may be used accordingly.

2) There are two sets of data, each of four classes, and these are treated entirely separately. These two may either confirm, or contradict, one another.

3) The Initial Ability factor was complex, having an attitude and an ability component and an interaction.

4) Three differences, instead of only 1, were evaluated for class differences.

5) Three interactions between classes and pre-abilities were evaluated.

6) Two Outcomes, change in attitude, and computing success, were evaluated.

7) The Analysis was by Computer, but using a regression approach similar to that used with the Synthetic data.

Some details of the predictors used, and of results are given in Table II and Table III.

Table II refers to the lower ability group.

Table III refers to the higher ability group.

The analysis was sequential, first with the 1st. predictor only, then with the first and 2nd., then with the 1st.,2nd. & 3rd. and so on. At appropriate stages the coefficients of the current multiple regression equation were printed, together with the loss in fit (sum-of-squares)due to removal of each one of the current predictors.

In the third last row of Table II we see that the regression equation is -

$(y-\bar{y}) = .0107(V_1-\bar{V}_1) +.0567(V_2-\bar{V}_2) +1.3378(V_3-\bar{V}_3) -.6053(V_4-\bar{V}_4) +.0095(V_5-\bar{V}_5)$

where -

$(y-\bar{y})$ is the deviation of Computing Success from the mean of all such.
$(V_1-\bar{V}_1)$ " " Pre-attitude to Maths. " " " "
$(V_2-\bar{V}_2)$ " " Post-achievement Maths. " " " "
$(V_3-V_3)$ " " the Class score $V_3$ " " " "
$(V_4-V_4)$ " " " " $V_4$ " " " "
$(V_5-V_5)$ " " " " $V_5$ " " " "

The Residual Sum of Squares - 90 degrees of freedom - is 828.822

The increase in Residual Sum-of-Squares on omitting any one variable is -

| | |
|---|---|
| Omission of $V_1$ only, | 2.06 |
| " " $V_2$ " | 35.98 |
| " " $V_3$ " | 474.50 |

This suggests that $V_3$ is the most important, and $V_1$ the least

On however, introducing the variables one at a time, in order, the reduction in residual sum-of-squares, calculated by Cochran's method(3), was

| | |
|---|---|
| on introducing $V_1$ | 42.62 |
| " " $V_2$ after $V_1$ | 1.43 |
| " " $V_3$ after $V_1$ & $V_2$ | 476.18 |

## TABLE IV

### Class / Teacher / Curriculum Effects from Tables II & III.

| | Lower Group | | Higher Group | |
|---|---|---|---|---|
| **Outcome = Attitude Rise** | M.S. | Sig. | M.S. | Sig. |
| Residual d/f | 86 | | 105 | |
| " M.S | 227.9 | | 97.8 | |
| $V_3$ | 3050.2 | *** | 8.8 | |
| $V_4$ | 202.6 | | 463.8 | * |
| $V_5$ | 1190.3 | * | 1039.9 | ** |
| Mean of 3 class differences | 1481.0 | *** | 504.2 | ** |
| **Outcome = Computing Achievement** | | | | |
| Residual M.S | 9.0 | | 14.7 | |
| $V_3$ | 476.1 | *** | 145.4 | ** |
| $V_4$ | 16.4 | | 65.1 | * |
| $V_5$ | 0.0 | | 10.7 | |
| Mean Of 3 Class Differences | 164.2 | *** | 73.7 | ** |

Probability of Occurance, solely by Chance - * less than 5%
** " " 1%
*** " " 0.1%

## TABLE V

### "Ordinary" Class / Teacher / Curriculum Variation.

| | Lower Group | | Higher Group | |
|---|---|---|---|---|
| | M.S. | Sig. | M.S. | Sig. |
| Outcome = Attitude Rise | 696.4 | P=6% | 751.8 | *** |
| " = Computing Achievement | 8.2 | n.s. | 37.9 | P=7% |

residual variance, allowing for uncontrolled variation in Teachers, of about $5\sigma^2$ where $\sigma^2$ is the variance due to errors in marking and in pre-testing. These errors of scholar indeterminance are averaged over a whole class of about 20-30 scholars.

To get reasonable precision we need to use more teachers and so reduce the errors in sampling the teacher population. That is to say, we need to exercise proper (statistical) care in equating, and sampling teachers. It will be argued that this can't be done in practice. If this is true, it is also true that one can't tell if the new curriculum is a general improvement! This seems sometimes to have been true; as Omar Khyaam found he -

> ".. evermore
> came out by that same door
> wherein I went!"

Fortunately the door is now open - by courtesy of statistics and multi-variate analysis. The door is a bit stiff, and we need to improve our use of those techniques of experimental design and analysis which have already been worked out.

### II (e) Notes on Analysis of Variance.

In the Analysis of Variance one calculates a number of variance components. For example, one component $V_4$ relates to the difference between results for classes II and IV, classes which are in the same ability grouping, and have the same teacher. The observed

These apparent contradictions as to the relative importance of $V_1$ and $V_2$ are confusing. The statements are correct. The situation is common. It occurs because $V_1$ & $V_2$ are correlated. Multiple regression analysis is designed to deal correctly with such situations, but it is often difficult to see what educational action is required. In such cases one might replace $V_1$ & $V_2$ by -

$P_1 = V_1 + V_2$

$P_2 = V_1 - V_2$

$P_1$ & $P_2$ will now contain all the information in $V_1$ & $V_2$, but most will generally be in $P_1$. This device is still fairly easy to understand, i.e. $P_1$ is akin to the average of all the predictors measured. Use of Principal Components of the predictors will give uncorrelated prediction, but their composition is difficult to visualise or understand. Also it will not be the same for the Lower and Higher classes. A compromise is usually desirable.

In planning one must arrange to achieve two results -

1) Find the effect, and significance, of the predictors of prime interest and importance, having first removed the effects of other predictors liable to confuse the effects of those of prime interest. Here the class, teacher, curriculum effects $V_3$,$V_4$,& $V_5$ were the prime interest, $V_1$ & $V_2$ being interferences due to differences in scholar abilities.

2) Obtain a realistic estimate of the error to which the effects in (1) are subject. This error must not be inflated by inclusion of assignable, but irrelevant effects.

In this case predictors $V_6$ - $V_9$ were introduced after the important predictors $V_1$ - $V_5$, and contributed little of significance. For significance testing, the residual after $V_9$ was used, but the small effect of $V_6$ - $V_9$ is some confirmation that little extraneous variation has been left in the residual.

## II (d) Discussion of Results - Real Data.

The residual mean squares are quite different for the two outcome variables - Rise in Attitude, and Computing success. This could well be due to differences in the scale of measurement, and is no surprise. However the residual mean squares also differ for the lower and the higher ability groups, and this was not anticipated. Presumably the tests used were similar. With higher ability one might expect greater individual initiative and more stable attitudes - who knows? There is also the possibility that the regression lines have different slopes for students in different classes. These possibilities have not been followed up.

The results we have are summarised in Table IV where significance levels are indicated by *'s. There is no doubt that significant differences are the rule, not the exception, between for both Post-attitude and for Computing success, and also in both the Higher and Lower ability groups.

In the lower group - Table II - the achievement of both class and teacher (I), using a special program of their own devising, is outstanding, both in achievement and in attitude rise. Circumstances are reminiscent of the Hawthorne experiment.

In each of the $V_5$ comparisons there were reasons for expecting superior performance, and these were known at the outset. In the other 4 comparisons there were no specific reasons for expecting the result to go one way or the other. The results in these cases of 'normal' variation are summarised in Table V. The results are not clear cut. The $V_6$ comparison is the discrepant one, an unusually high discrepancy for attitudes, and a low one for achievement.

The general feeling is that differences between classes are likely to occur in the ordinary school operations of allocating students to classes, and that the residue of these differences - after allowing for pre-measurement of attributes - is likely to be significant.

A typical value of the basic Variance-ratio looks like about 4 to 5, but rises considerably in cases where one can pin-point a difference beforehand. If we wish seriously to experiment with different curricula, or different teaching methods, we must reckon on a difference indicates the presence of variation somewhere. The unit on which measurements are made is the individual scholar, so one calculates the random variance between the scores of individual scholars which would be likely to give rise to the observed difference between class totals. By inverting the process, the observed difference becomes an estimate of the variance between individual scholars. If the variance between scholars so calculated is far higher than is known, on other evidence, to be the case, then one concludes that the result is a sign of (or significant of) some cause or component of variation not included in the other evidence.

At this point it is often helpful to remind ourselves that the variance of a population is the average of the squares of the deviations of individual observations from the population mean $\mu$. More specifically, if we are discussing the difference between two nominally similar classes, the long term average difference is zero. If the observed difference is $d_1$ then $d_1-\mu$ is $d_1$, and the variance estimate is $d_1^2$. If, as in our case the class size is about 24, and if the variance of the population of students from which classes are selected is $\sigma^2$, then the variance of the mean of a class - a sample of 24 - is $\sigma^2/24$. The variance of the difference of means of two such classes is $2\sigma^2/24$, or $\sigma^2/12$ and this is revealed as $d_2^2$ where $d_2$ is the difference of class means. Hence we estimate $\sigma^2$ from 'between class' averages as -

$$d_2^2 = \sigma^2/12$$

or

$$\sigma^2 = 12d_2^2$$

If we work with $\underline{D}$ = difference of class totals, instead of with $\underline{d}_2$= difference of class means, then $\underline{D} = 24d_2$, $D^2 = 576d_2^2$ and

$$\sigma^2 = D^2/48$$

This assumes there are no 'assignable causes' of difference between classes other than the variance $\sigma^2$ within classes. A significant result would then indicate the presence of an assignable cause.

We know there are assignable causes. Allocation to class is by interest, and by expected future curriculum. We also know that a class.teacher combination tends to become a social entity in its own right. We recognise these things, but often dont attempt to quantify their elements.

We also recognise that within each class there 24 individually different scholars. The school has individual 'ability' test results from which we may infer approximately, and allow for, these individual differences. This has been done. This also, to some extent, allows for the consequential differences between classes. What we are talking about as $\sigma^2$ is therefore the 'residual variance' after allowing for that part of individual differences which can be inferred from the schools ability testing. Most of this residue is found 'within' classes, but some of it appears in the variance 'between' classes. Our data shows the latter to be larger than the former which means that some components of the residual variance are associated more with 'between' than with 'within' class comparisons. One of the factors which we recognise as having this charistic is 'the effect of teacher on class performance'.

The standard deviation of the difference between class means is -

$$\pm\sqrt{(\sigma^2/12)}$$

In addition to a difference of about this size, we also have a non-random difference $\underline{f}$ related to teacher differences, and together giving a difference between means of about

$$f\pm\sqrt{\sigma^2/12}$$

Expressed as a variance, this becomes approximately

$$f^2+\sigma^2/12$$

This expression appears to have a value about 5 times $\sigma^2/12$, hence

$$f^2 = 4\sigma^2/12$$

$$f = 0.6\sigma$$

Our data indicate that significant class differences exist, and that when there are no causes that we specifically identify, there will, in fact, be a teacher effect of about the size of the standard deviation of individual scholar deviations after allowing for pre-tests. Put another way, by using a large class we reduce error in class mean to about 0.2σ but we retain a teacher error of 0.6σ and graft this deviation onto each of 24 students.

This is a rough indication, but not rough enough to ignore.

## III CASE II.

### (a) Introduction.

The second example from my case book will be dealt with rather briefly.

The School was rather large, but not prestigeous. Again there was interest in comparing the effectiveness of different curricula, and in looking at the merits of 'streaming, in the context of low ability of some scholars.

Initial thinking was dominated by the magic number 3.

There were 3 terms.

There were 3 established curricula which might be compared.

There were 3 teachers available.

There were 3 class spectrums of interest -

i) High / Medium ability.
ii) High / Medium / Low ability.
iii) Medium / Low ability, i.e. relatively un-selected.

It was known that experimental designs called 3 x 3 Graeco-Latin squares existed in which 4 independent 3-level factors could be used with 9 experimental units without confounding main effects.(6,7,5) Initial planning was based on these designs.

As the starting day approached, things began to go wrong.

Material for one of the Curricula could not be obtained

Some scholars were drafted out to a special preparatory course.

The upshot was a change of design to 6 classes or experimental units, with 2 curricula, but the original 3 class spectrums. A scholar was taught by only one curriculum, but in the course of the year was taught by all 3 teachers, and each teacher taught all scholars. An outline design is given in Table VII.

TABLE VII

3 x 2 x 3 Design for Class Spectral types, Curricula, and Teachers.
Three Terms or Blocks

| | | Term 1 | | Term 2 | | Term 3 | |
|---|---|---|---|---|---|---|---|
| Curriculum | | X | Y | X | Y | X | Y |
| Class | M/L | A | B | C | A | B | C |
| Spectrum | H/M/L | C | A | B | C | A | B |
| | H/M | B | C | A | B | C | A |

This experimental design confounds teacher differences with interactions of the other 3 factors. Analysis was carried out by computer using Yates(6) tabular method as modified by Bainbridge(8) and the confounding was sorted out by feeding the errors through the same process. See Bainbridge(9) and Hartley(10). Table VIII shows the confounding. The table is arranged using Class spectrum, Curriculum, and Term as factors, and Teachers as Errors. The class types are coded 0, 1, 2 in the order given above; the curricula are coded 0, 1 and the terms 0,1,2. This gives the table in the standard notation of Yates. Where numbers appear in the right-hand column, e.g. 2,-1,-1 under A,B,C at the head of that column, it means that twice the difference between A and the average of B and C is added to the effect indicated in the 3 leftmost columns and the same row viz. 1 0 1 which, in the standard notation is the change in the excess of class-type H/M over M/L on going from term 1 to term 3. Making reasonable assumptions about where the realities, and where the rubbish lie, one can unscramble a lot of the Teacher differences.

This design with 6 classes and 3 teachers is considerably more informative than is the Case I design with 8 classes and 6 teachers - but for other difficulties.

## TABLE VIII

### Case II Confounding for Table VII

| Factorial Code sct | Class Spectrum | Curriculum | Term | Teacher. | Class Spectrum | | | Curriculum | | | Term | | | Divisor. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | A | B | C | A | B | C | A | B | C | |
| 000 | M/L | X | 1 | A | 1 | 1 | 1 | 2 | 2 | 2 | 6 | 6 | 6 | 18 |
| 100 | H/M/L | X | 1 | C | 1 | 1 | 1 | 2 | 2 | 2 | . | . | . | 12 |
| 200 | H/M | X | 1 | B | 1 | 1 | 1 | 2 | 2 | 2 | . | . | . | 36 |
| 010 | M/L | Y | 1 | B | 1 | 1 | 1 | -1 | . | 1 | . | . | . | 18 |
| 110 | H/M/L | Y | 1 | A | 1 | 1 | 1 | . | 1 | -1 | . | . | . | 12 |
| 210 | H/M | Y | 1 | C | 1 | 1 | 1 | 1 | -1 | . | . | . | . | 36 |
| 001 | M/L | X | 2 | C | -1 | 1 | . | -1 | 2 | -1 | . | . | . | 12 |
| 101 | H/M/L | X | 2 | B | . | -1 | 1 | 2 | -1 | -1 | 2 | -1 | -1 | 8 |
| 201 | H/M | X | 2 | A | 1 | . | -1 | -1 | -1 | 2 | . | -3 | 3 | 24 |
| 011 | M/L | Y | 2 | A | -1 | 1 | . | . | . | . | . | . | . | 12 |
| 111 | H/M/L | Y | 2 | C | . | -1 | 1 | . | . | . | . | 3 | -3 | 8 |
| 211 | H/M | Y | 2 | B | 1 | . | -1 | . | . | . | 6 | -3 | -3 | 24 |
| 002 | M/L | X | 3 | B | 1 | 1 | -2 | 1 | -2 | 1 | . | . | . | 36 |
| 102 | H/M/L | X | 3 | A | -2 | 1 | 1 | -2 | 1 | 1 | . | -3 | 3 | 24 |
| 202 | H/M | X | 3 | C | 1 | -2 | 1 | 1 | 1 | -2 | -6 | 3 | 3 | 72 |
| 012 | M/L | Y | 3 | C | 1 | 1 | -2 | -3 | . | 3 | . | . | . | 36 |
| 112 | H/M/L | Y | 3 | B | -2 | 1 | 1 | . | 3 | -3 | 6 | -3 | -3 | 24 |
| 212 | H/M | Y | 3 | A | 1 | -2 | 1 | 3 | -3 | . | . | -9 | 9 | 72 |

The Effects are:- For Class Spectrum S1 (-M/L +H/M)
S2 (M/L-2(H/M/L)+H/M)
For Curriculum C1 (-X + Y)
For Term T1 (-(1) +(3))
T2 (+(1) -2(2) +(3))

It is necessary as before to remove the irrelevant assignable causes of variation which otherwise mask significant effects. The method used in this case was as follows -

A fact of life was that a number of students dropped out, leaving uneven classes. By random selection each class was reduced to 24 candidates, who completed the year and were used in subsequent analysis.

The initial selection of class members for the three profiles of class ability, was based on some quick ability and interest tests at the time of entry. These test results were used to divide the final 24 scholars into four ability quartiles, each of 6 scholars - the Top, 2nd.,3rd, and Bottom quartiles, which provided another factor for the Factorial Analysis table. The variation within quartiles was calculated as Residual Variance. The variation between quartile totals was obtained from the standard Orthogonal Polynomial Functions(7) for factors at 4 levels. These are -

| Quartiles | | Top | 2nd. | 3rd. | Bottom | Divisor. |
|---|---|---|---|---|---|---|
| Function | 1 | 3 | 1 | -1 | 3 | 20 |
| " | 2 | 1 | -1 | -1 | 1 | 4 |
| " | 3 | 1 | -3 | 3 | -1 | 20 |

Function 1, linear with pre-ability, showed significant results, but not functions 2 or 3.

The experiment got through the first year. With only 6 classes there were not enough degrees of freedom to show reliably the presence of assignable causes of error. For example the three classes on one curriculum should form a progression - hopefully, but not necessarily a linear one. If the non-linearity is not the same on the 2 curricula, then some error has occurred in class selection or treatment. This must be included in the error to which curriculum and class profile comparisons are subject. There were 3 such. In addition 6 other errors were extracted when correcting for the confounding caused by teacher differences. The analysis showed significant teacher and curriculum effects, and the suggestion that, on the whole, interactions of these factors were not negligable.

It was decided to repeat the experiment in the following year.

The school was not prestigeous! Next year there remained only one of the three teachers engaged on the experimental teaching. The experience gained had been lost. Two new teachers were appointed for the work.

Without justification, one may suppose another situation. It is not easy to think ones way through any analysis such as this. Most teachers couldnt, or wouldnt bother to, follow it. The natural reaction is not to say it is too hard, but to say it is rubbish! It would appear that the two new teachers became infected with this atmosphere. At any rate, in the second year,the continueing teacher was far better than either of the new teachers. They in turn produced very erratic results. It became necessary to study the confounding of Teacher x Curriculum interactions as well as those shown in TableVIII. It seemed that these two teachers were not attempting to teach their classes according to the documented curriculum specified for their classes.

We meet a new idea. School morale is important! Have I heard that elsewhere?

## IV CONCLUSION.

### Consequences to Educational Administration of Teacher Differences in Curriculum Studies and in Studies in Institutional Differences.

Teacher differences, Class differences, and Institutional differences need to be taken into consideration in planning experiments on alternate curricula. With an error variance 5 times higher than expected, we need a large, efficient, and well planned experiment with a large team of teachers and covering a wider range than has often been usual. We must expect Curriculum x Teacher interactions. If the same curriculum is the best for all Institutions, there is no need for different sorts of Institution. We see then the need for large experiments involving teaching in a number of schools having some commonality in their aims, but also quite distinct differences. We then have the opportunity to see which methods work better, and which worse, in the several environments. We cam also move toward assessing the standards, and performance, of schools in different educational areas. In addition such a widely based use of evolving curricula by many teachers provides excellent de-bugging, and is also the best preliminary way of generating user acceptance. After all, a better method is useless unless it becomes accepted and used.

It is generally believed that there are important socio-economic differences between schools in different areas, and such experimenting lends itself to assessment of these differences, and to distinguishing them from other causes of differing performance.

The organisation of such studies needs good statistics and good cooperation. There must be an organiser which is an'outsider'to most,or all, schools.

For the Educational Administrator, creation of men with the personalities and experimental skills to plan and coordinate such projects - and creation of an organisation which welcomes the discovery, and application of better methods - become the twin keys to progress.

**TABLES.**

I Case I Classes, Abilities, Teachers. page 1

II Case I Regression Analysis; Lower Ability Group page 5

III Case I " " Higher Ability Group page 6

IV Case I Class / Teacher / Curriculum Effects from Tables II & III. page 8

V Case I Ordinary Class / Teacher / Curriculum variation. page 8

VI Omitted

VII Case II 3 x 2 x 3 Design for Class spectral types / Curricula / Teachers and 3 Terms or Blocks page 11

VIII Confounding for Table VII page 12

FIGURE I Case I Synthetic Data. Regression Lines. page 4

**REFERENCES.**

1) Yule,G.U, & Kendall,M.G, "Introduction to the Theory of Statistics" 1911-58 Charles Griffin, London

2) Fisher,R.A. "Statistical Methods for Research Workers" Oliver & Boyd, Edinburgh.

3) Cochran,W.G. "The Omission and Addition of an Independent Variable ..." J.Roy.Stat.Soc. Supp. 5, 1938, p171.

4) Davies,O.L, "Statistical Methods in Research & Production" Oliver & Boyd, Edinburgh. 1958, 3rd. Ed.1967.

5) Cochran,W.G. & Cox,Gertrude,"Experimental Designs" Wiley, New York 2nd. Ed. 1957.

6) Yates,F, "Design and Analysis of Factorial Experiments" Commonwealth Agricultural Bureaux, Farnham Royal, Bucks. 1937-1964.

7) Fisher.R.A, & Yates,F. Statistical Tables. Oliver & Boyd, Edinburgh.1938-49

8) Bainbridge,J.R,Grant,Alison, & Radok,U. J.Amer.Stat.Assn., Mar. 1956,V51 pp149-158.

9) Bainbridge,J.R, "Factorial Experiments in Pilot Plant Studies" Ind.Eng.Chem. 1951, p1300.

10) Hartley,H.O. Analysing Errors & Data by the same process to show effects of the former. Lecture to Int.Biomet.Soc. 1963, No exact ref.

# AN APPROACH TO EVALUATING TERTIARY COURSES

*William C. Hall*

## INTRODUCTION

This paper describes an approach to course evaluation developed by the Advisory Centre for University Education since its formation four years ago. During that time we have conducted 18 full-scale course evaluations.

An important feature of this work is that evaluations are conducted with (rather than for) a department. Although all results of evaluations are confidential, some of the people with whom we have had contact have kindly agreed to allow us to describe part of this work with them.

This paper is certainly not intended to describe the approach to evaluation! It is presented as one approach that has been found to be successful when the evaluator is subjected to very definite practical constraints (especially the constraint of time).

Most of the paper has been produced in an A.C.U.E. publication "Evaluating Courses" obtainable from the University of Adelaide.

## A DEFINITION OF EVALUATION

The definition of evaluation used in this paper is *seeking to obtain answers to questions that will enable decisions to be taken about courses*. This definition makes the following assumptions:

(a) meaningful *questions* can be asked about courses;
(b) *techniques* are available to provide information that will enable useful answers to these questions to be given;
(c) the right *people* are available to administer the techniques and there are other people who are willing to have the techniques applied to them;
(d) *decision-making* is an essential part of course construction and development.

There are 5 stages in the approach to course evaluation which has been developed:

STAGE 1 Determining the questions that need answering about a course.

| | |
|---|---|
| STAGE 2 | Producing an evaluation schedule which is discussed at a departmental meeting (or with an individual, if only one person is running the course). |
| STAGE 3 | Applying techniques to obrain preliminary answers to questions. |
| STAGE 4 | Applying further techniques to obtain more specific answers to questions. |
| STAGE 5 | Writing a confidential, descriptive report of the evaluation for the department. This report is discussed at a departmental meeting, after which our copies are destroyed. |

## REASONS FOR EVALUATING

There are three main reasons why course evaluation should be undertaken in tertiary institutions. Firstly, an evaluation of *input* (e.g. technical resources, the learning environment, staff capabilities, course administration, student expectations, students' prior knowledge) can reveal deficiencies in course planning. Secondly, an evaluation of *process* (e.g. course aims, course content, ways of teaching, methods of learning) can provide feedback for on-going course modification. Thirdly, an evaluation of *outcomes* can enable future changes to be made on the basis of solid information, can offer help to those who are thinking of adopting the course, and (a reason which is becoming increasingly important) can provide a statement of accountability.

It sometimes seems to me that evaluations are conducted at the wrong times. For example, in my first year at the University of Adelaide about 10,000 duplicated lecture evaluation questionnaires were administered during November (i.e. at the end of courses) whereas only a few hundred were used during the earlier parts of that year. This meant that students were not well motivated to complete most of the questionnaires (the results could not affect them) and staff are unable to pursue to a greater depth any of the gross generalisations revealed from the results of this type of general survey. *The time to start thinking about an evaluation strategy is when a course is first being planned.* Evaluation should not be a "seal of respectability" but a tool in decision-making.

## EVALUATION, ASSESSMENT AND MEASUREMENT

The terms "evaluation", "assessment" and "measurement" are sometimes used interchangably. They are not synonymous. The definitions used by me are as follows:

| | |
|---|---|
| EVALUATION | Seeking to obtain answers to questions that will enable decisions to be taken about courses. |
| ASSESSMENT | The application of techniques to discover how successfully students have achieved course aims. (The commonest technique at a university is probably the essay examination.) |
| MEASUREMENT | The application of objective investigation and scientific experimentation to education. (A common tool of measurement is the standardised test.) |

Assessment can be (and often is) one component of evaluation.

Measurement can occur in both assessment and evaluation (but rarely does).

My approach to course evaluation is not classical educational research. Control groups, for example, are not possible. To use the Campbell and Stanley jargon, my approach is closest to the pre-experimental or one-group pre-test, post-test design. This means that there are probably numerous sources of invalidity and so any results need to be interpreted with caution. One way to help overcome the validity problem is to apply a number of different techniques to answer the same question(s).

## COURSE EVALUATION

The five stages to course evaluation have already been listed. The following table summarises the evaluator's actions for every stage.

| STAGE | EVALUATOR'S POSSIBLE ACTIONS |
|---|---|
| 1. Determining the questions that need answering about a course. | Interviewing staff.<br>Interviewing students who have completed the course.<br>Reading departmental literature.<br>Direct observation. |
| 2. Evaluation schedule. | Departmental meeting. |
| 3. Applying techniques to obtain preliminary answers to questions. | Administering general questionnaires to students and/or to staff.<br>Discussions with groups of students and/or with staff.<br>Direct observation. |
| 4. Applying further techniques to obtain more specific answers. | Sound or video recordings.<br>Specific questionnaires administered. (Questionnaires at this stage are usually open-ended.)<br>Interviewing groups of carefully selected students.<br>"Living" with an average student.<br>Interviewing staff.<br>Direct observation. |
| 5. Descriptive report. | Departmental meeting. |

## STAGE 1 DETERMINING THE QUESTIONS THAT NEED ANSWERING

There are two parts to this stage: firstly, the important variables have to be identified (e.g. particular groups of staff, particular groups of students, aspects of the environment) and, secondly, the important questions to be answered for each of the variables have to be formulated.

A department might approach the evaluator with the suggestion that "they are interested in what students think about their methods of assessment". Two important variables would seem to be the students and the assessment procedures. Staff interviews might reveal that although the new assessment format appears to be reasonable, no account has been taken of assessment demands in other subjects being studied by students. Departmental literature reveals that there is no indication of how the proposed assessment fits in with the total curriculum (that is, the content of the course and the way it is taught do not seem to be connected with the assessment proposals).

Preliminary questions about the two variables (students and assessment) together with a third variable (staff) now start to emerge.

| | |
|---|---|
| STUDENTS | What workload is demanded by the new assessment scheme?<br>What, and when, are other assessments completed during the year?<br>Do students think that the assessment reflects the nature of the course? |
| STAFF | What workload is demanded by the new assessment scheme?<br>Do staff think that the assessment reflects the nature of the course? |
| ASSESSMENT | What is the specification matrix for the assessment scheme? (What aims are being assessed? and What are the component weightings?)<br>How do the aims and the weightings on the matrix compare with how the course is taught? |

Already the evaluation is starting to become both lengthy and complex. The next stage is to draw up an evaluation schedule.

The importance of direct observation of the course at this first stage cannot be over-emphasised. Attendance by the evaluator at every teaching session over a period of 1-2 weeks is essential.

When students are interviewed at this stage they are usually those who have already completed the course. A random 10-15% or 15-20 students, whichever is the larger, are chosen and they are interviewed in groups of 3-4 in an informal room. Three questions are generally asked:

A. Please will you describe the course?
B. What were the particularly good things you remember about the course?
C. What were the disappointing features of the course?

About 30-40 minutes are spent with each of the groups of 3-4 students.

Two consequences of the direct observations and interviews are, firstly, there will be modifications to the questions to be answered as a result of the evaluation and, secondly, constraints to the administering of evaluation techniques will start to emerge.

## STAGE 2 EVALUATION SCHEDULE

The evaluation schedule lists the variables and associated questions to be answered and shows what techniques will be applied to obtain preliminary answers to these questions. Suggested dates for administering the techniques are also given.

A hypothetical schedule for the questions obtained in the previously described Stage 1 now follows.

| Variable | Questions to answer | Proposed evaluation technique | Suggested date |
|---|---|---|---|
| Student | 1. What workload is demanded by the assessment. | Diary | At intervals throughout the year. |
| | 2. What and when are other assessments completed during the year? | Information from other departments. | When convenient. |
| | 3. Does the assessment reflect the nature of the course? | Questionnaire | Mid-term 3. |
| Staff | 1. What workload is demanded by the assessment? | Questionnaire | Immediately after assessment |
| | 2. Does the assessment reflect the nature of the course? | Interview | Term 3. |
| Assessment | 1. What is the specification matrix? | Departmental discussion. | Term 1 |
| | 2. How does the matrix compare with how the course is being taught? | Departmental discussion. | Mid-term 3. |

This schedule allows for feedback to staff and to students throughout the year (and enables any necessary modifications to be made to the evaluation itself). It also allows for more specific answers to the questions to be obtained (Stage 4) during the final part of Term 3.

## STAGE 3 APPLYING TECHNIQUES TO OBTAIN PRELIMINARY ANSWERS TO QUESTIONS

Examples of techniques used at this and at the next stage will be given.

## STAGE 4 APPLYING FURTHER TECHNIQUES

At this stage, questionnaires are usually open-ended and arise directly from the results of Stage 3. Interviews are conducted with a *carefully selected* group of students (the selection is based on academic performance, age, sex and previous courses taken). About 10-15% of the class or 15-20 students, depending on which is the larger, are interviewed in groups of 3-4.

Both sound and video recordings have been used at this stage of evaluations. For example, in an early study for the Department of Medicine, students were provided with sound cassettes. Recorders were switched on whenever a student interviewed a patient and the recordings were later analysed. Video-recordings were also used in the evaluation.

A successful, but time-consuming, technique is for the evaluator to "live" with an average student ("average" being based on academic performance) for a minimum period of 2 days. A diary of student activities can be kept by the evaluator. Also, by this stage the evaluator has particular, detailed, questions he is interested in pursuing.

## STAGE 5 THE DESCRIPTIVE REPORT

The final report summarises the evaluation process and describes the observations made. Generalisations, whenever possible, are produced but no recommendations are provided. These emerge at the departmental meeting when the report is being discussed.

The final report might include suggestions on, for example, how teaching or assessment could be varied. One example of this is an assessment schedule that was produced as a consequence of an evaluation of clinical skills in Restorative Dentistry.

After the report has been discussed the A.C.U.E. may be involved in dealing with changes to courses that are necessary as a result of the evaluation. Some of these changes will be described in the presentation.

# Curriculum Problematics in a Poly-ethnic Context

(Paper presented to the Conference of The Australian Association for Research in Education. Canberra, November 11-13, 1977)

B.M. Bullivant
Monash University

## Introduction

The emergence of various perspectives on the curriculum has been a feature of the seventies, and has taken forward the earlier thinking of such theorists as Taba (1962) and Wheeler (1967). The publication of Michael Young's Knowledge and Control (1971) in itself provided a number of key ideas. The curriculum is a means of controlling power, and 'processes' both knowledge and pupils (Bernstein, 1971). It should, and does indeed, reflect the values of society and its culture (Bourdieu, 1971). It is also intimately related to the types of espistemologies and logics operating within groups and between groups (Keddie, 1971; Horton, 1971 ). The same notion has been brought out in Keddie's later publication (1973) in which the culturally constructed nature of knowledge is stressed (see also Frake, 1973; Gladwin, 1973). The concept of culture has also been given prominence in the recent writings of such theorists as Lawton (1975) and Reynolds and Skilbeck (1976). The latter take their direction for this from a seminal, early work of Smith, Stanley and Shores (1957).

In view of such a valuable ferment of ideas it might appear to be unduly carping to observe that the majority of them take little if any account of the culturally pluralist nature of many Western societies, except in terms of class differences which are still of major concern. Their ethnic composition is now being recognized in the United States (Banks, 1975; Banks et al., 1976) in a general social movement known as the 'new pluralism' or 'ethnic revitalization'. Its essential characteristic has been stated by Glaser and Moynihan (1975: 2-3): Ethnicity is 'a new social category ... [marked by] a pronounced and sudden increase in the tendencies by people in many countries and in many circumstances to insist on the significance of their group distinctiveness and identity and on new rights that derive from this group character'. This paper addresses itself to the problems of the curriculum that flow from a society having these characteristics, as a contribution towards the steadily growing corpus in anthropology of education. It is intended as an exploratory exercise in curriculum meta-theory, i.e. an attempt to make theoretical speculations about theories current in the area of the curriculum.

## Curriculum as communication system

The definition of the curriculum proposed here is a refinement of earlier models used to analyse schooling and education from an anthropological perspective (c.f. Bullivant 1972, 1973a, 1973b). The curriculum is a selection from a socio-cultural group's stock of valued traditional and current public knowledge and experiences, usually purposefully organized by such educational agencies as schools, and made available to pupils sequentially through syllabuses and units.

The curriculum occupies a strategic place in the total process of enculturation by which the social group's culture is transmitted to each new generation. This must occur if the group is to perpetuate itself and survive as an on-going entity, i.e. enculturation is a necessary condition for societal survival. However, this process need not necessarily be accomplished through schooling, hence the force of the word 'usually' in the definition of the curriculum. Its inclusion is designed to alert us to the fact that some pre-literate communities do not have schooling in the accepted sense, yet still manage to transmit their cultures quite satisfactorily. Despite this, it would still be quite appropriate to adopt the term curriculum for what is taught to the young. For instance, the formal enculturation of adolescent Masai boys the moran, in Kenya takes place in special forest camps in which they live during adolescence and undergo a series of rites de passage (the 'curriculum'). To call these schools in our sense would be a misnomer. Similarly Australian Aboriginal communities such as the Walbiri still teach young boys tribal knowledge in men's groups, but, again these could hardly be described as schools in our sense.

The universality of the curriculum rests on the meaning of culture adopted here. It is defined as a patterned system of symbolically and extra-symbolically communicated interdependent knowledge and conceptions about the technology and skills, customary behaviours, values, beliefs and attitudes a society has evolved from the past, and progressively modifies to give meaning to and cope with the present and anticipated future problems of its existence. Curriculum can now be seen as a selection of that knowledge and conceptions.

The definition is an attempt to stress the technical sense of culture as a social group's problem-solving, survival device evolved over time to cope with the challenges posed by three conceptually separable environments. The first is the natural or geographical environment and its biotic resources which part of the group's technology and associated skills are designed to exploit and master. The second is the social environment comprising other groups of people and the social relations of the group itself. Relationships with this environment involve using patterned customary behaviours with their associated norms and sanctions. Their style and strength derive from the values members of the group place on social relations and the attitudes they hold towards each other and other groups. The third environment is more speculative and can be termed the 'metaphysical environment'. This comprises phenomena and occurrences in nature for which we have no scientific nor rational explanations, e.g. supernatural forces, the dramas of birth, existence, death and the hereafter, natural disasters and powers in the universe that, to some, appear to defy explanation. To give meaning to these problems of existence social groups devise belief systems such as religions, animism, even astrology, numerology and the currently fashionable occult craze. All these are designed to provide 'answers' or the 'meaning of it all'.

The definition of culture is framed in such a way as to bring out a feature stressed by modern structural anthropologists of the Levi-Strauss school that culture can be communicated both symbolically and extra-symbolically, i.e. by means that lie outside the realm of symbol. Leach has suggested that 'Culture communicates; the complex interconnectedness of cultural events itself conveys information to those who participate in those events' (Leach, 1976:2). Indeed, to this theorist, all aspects of culture communicate something. I take a similar position in holding that all aspects of a school communicate cultural information (Bullivant, 1975:385 ff.): the actions of persons working in it, the arrangement of buildings, display material, artifacts of all kinds, use of space such as playing fields, gymnasium, school chapel, and so on.

The universality and central place of the curriculum in the enculturation process justify considering it as an institution comparable to such institutions as the economy, law, family, and education. Each of these is 'an interrelated system of social roles and norms organized about the satisfaction of an important need or function' (Theodorson & Theodorson, 1970:270). In this case the curriculum has the major function of acting as a communication system for the transmission of a group's culture. Its direction and details of operation derive from a 'charter' in the Malinowski sense (1944) of a body of beliefs, ideas, attitudes and values held by the group which determine the ends and means of the institution's function.

For example, in the Jewish ethnic day school and its supporting community studied by the writer (Bullivant, 1975) the religious curriculum clearly derived from a body of prescriptions set out in the Pentateuch and elaborated in the Mishnah (Talmud). Their influence could be seen in the content, teaching style, and other aspects of the religious studies programme. The ends in view were quite explicit: to produce observant Jews in obedience to an injunction, that was held to be of Divine origin.

'And thou shalt teach them [the Commandments] diligently unto thy children' (Deut. 6: 7). 'And all thy children shall be taught of the Lord and great shall be the peace of thy children' (Is. 54: 13).

Figure 1 about here

As the above diagram illustrates, the charter also influences other components of the institutions. These are the personnel involved in its operation, their behaviours, role expectations and governing norms, together with the apparatus and equipment they use. All these will be present in any institution, but their precise nature will vary from culture to culture. Thus, although the curriculum of the Walbiri initiation group will be markedly different from the Western school also attended by Walbiri children, all the conceptual components that make up an institution can still be distinguished for analysis. For the argument that follows it is vital to appreciate this point. All socio-cultural groups have similar institutions, as they all have similar needs. But the way the institution is organized to carry out its functions will vary from culture to culture. When we recall that 'all culture communicates', the curriculum as communication system in one socio-cultural group (A) will convey different cultural realities through different message components compared with the curriculum of a socio-cultural group (B). In the final analysis, a child from (A) being taught through the curriculum appropriate to (B) may not 'get the message' for reasons that make the communication system problematical. These types of curriculum problematics will be taken up at a later stage of this paper.

## Curriculum as power system

When we narrow the focus of analysis to consider the Western-type school, we are in effect looking at one way of institutionalizing the curriculum, and each component of the school can be related to the diagram of the institution above. But it is possible to conceptualize the curriculum in a different way as a power system. This aspect is implicit in the definition of the curriculum, which, we recall, involves selecting, organizing, and making knowledge available (implementing it) to pupils. These can be thought of as various levels of the curriculum as shown in the following diagram.

Figure 2 about here

Each level involves making value judgements and choices between a variety of competing aspects. At the level of selection a choice has to be made from the vast amount of knowledge and experiences available in any particular culture. A philosophical decision has to be made about the kind of knowledge that is most worthwhile to put in the curriculum and what to omit. This will be determined in part by an 'ends question' which forces us to consider what the child's schooling is for. If we see it as being to prepare him or her to work in and for society a different selection of 'work-oriented' knowledge will be needed compared with that which might be selected to develop the child as a complete individual — 'person-oriented' knowledge. This in turn makes us consider the type of society we have at present (a sociological question) and the type of society we want (a socio-philosophical question).

At the level of organization the educator has to make value-laden decisions about the allocation of scarce resources such as time-tabling, staff, rooms, equipment, materials and all the school paraphernalia we are familiar with. None of these is in unlimited supply, and is thus subject to the economic laws of resource allocation. The organizational level also involves making decisions about sequencing and grouping children's learning. These in turn necessitate deciding which psychological learning theories ought to prevail in the curriculum. Such relatively simple aspects as automatic promotion from grade to grade versus delayed promotion based on mastery learning, or 'streaming' versus 'setting', depend on the view one takes about the learning process.

At the implementation level similar decisions have to be made, but they are further complicated by the problems each teacher faces in having to devise the most effective learning-stimulus mix in his classroom. This is a form of interaction setting with both animate and inanimate components (Bullivant, 1975:372 ff.), in the form of the actors present with the architectural design and artifacts which provide the spatial context for the interactions. They all generate interaction incidents. These can take place between those persons present, between any one of them and his 'self' (i.e. self-interaction), between an actor and an invisible Other, such as was observed by the writer when boys were praying in the synagogue, and between an actor and an inanimate component in the interaction setting. In the classrooms, for instance, pupils 'interact with' their desks, burrowing around for books and so on. To achieve any complete description of the classroom this type of interaction incident has to be taken into account.

Teachers can arrange the learning-stimulus mix in a virtually endless number of ways. To take just two examples, we can have open classrooms versus traditional classrooms in which the desks are arranged in rows facing the front of the room. Display boards on the walls can be utilized to show a wide range of materials with which pupils are encouraged to interact at will, or they can be free of materials with the major visual stimuli coming from handouts provided by the teacher at the beginning of the lesson.

At least two aspects of the three levels of the curriculum must be stressed. Firstly, the way they are handled in the curriculum will reflect its socio-cultural group's preferred cultural style and broad values. For example, a group that values discipline and strong control over children is unlikely to encourage a form of curriculum implementation that allows an open classroom learning-stimulus mix. The same group might see 'received knowledge' as being the most important as in Amish or Hutterite communitarian education (Hostetler, 1970; Hostetler & Huntingdon, 1971), and only allow a strong hierarchical form of teaching to take place. To another group with strong liberal, democratic values, the child is allowed to pursue a self-directed form of discovery learning with little overt teacher direction. The comparisons between education in the U.S.A. and U.S.S.R. made by Bronfenbrenner (1974) are excellent examples of the cultural influences on the curriculum. A similar approach has been used by Dahlke (1958). In short, the style of the curriculum in all its aspects will communicate the culture.

But for the purposes of this analysis there is a second underlying theme in the levels of the curriculum. As this writer has shown in an extended treatment (Bullivant, 1978a) they are all vehicles in the school for exercising various forms of power. This is defined as 'an actor's ability to induce or influence another actor to carry out his directives or any other norms he supports' (Etzioni, 1969:60-2). There are three major forms of power. 'Coercive power rests on the application, or the threat of application, of physical sanctions ... generation of frustration through restriction of movement; or controlling through force the satisfaction of needs'. Coercive power operates most obviously in the interaction setting of the classroom in relations between teacher and taught. 'Remunerative power is based on control over material resources and rewards'. We can see this form of power operating at all levels of the curriculum but particularly at the organizational level. 'Normative power rests on the allocation and manipulation of symbolic rewards and deprivations'. This form mostly operates at the implementation level.

In a monocultural society, the control over power through the 'classification' and 'framing' of knowledge in the curriculum operates to advance the interests of social classes and dominant groups in society as Bernstein has shown (1971). Their effect is both overt and covert, and, can even be shown to operate in the field of curriculum development where a form of 'hidden curriculum' is at work to perpetuate middle-class, academic control over curriculum innovations, despite research workers' assumptions that they are catering for the needs and interests of children from disadvantaged backgrounds (Bullivant, 1973a). In a culturally pluralist society like Australia, such class-focussed analysis must be augmented to take into account cultural and ethnic factors. These transcend class and inject a new dimension into our thinking, viz., the perpetuation of ethnic hegemony by one socio-cultural group over others. This is achieved through both features of the curriculum — communication and power.

## The poly-ethnic society

A full treatment of the operation of ethnic hegemony and the social realist theories on which it is based is beyond the scope of this paper, but is considered in detail elsewhere (Bullivant, 1978b). The present discussion rests on an analysis of the nature of Australian society. This is an indispensable pre-requisite for diagnosing and prescribing for the needs of its education system in general and the curriculum in particular.

Demographically, the Australian population aggregate is made up of persons from a wide variety of national origins. Approximately twenty five percent of the population were either born overseas or are children of persons born overseas. The variety is immense. In Victoria, as an example, some one hundred national and language groups have been identified among the immigrant sector of the total population alone. Using another criterion, Australian society is composed of a great variety of socio-cultural groups. Besides national origins, these can be distinguished according to religious affiliation, language, socio-economic 'class', maintenance of distinctive customs and historical traditions. In addition, there is a proportion of the population whose roots and traditions are of far more ancient origin. These are the Australian Aborigines and Torres Strait Islanders of various degrees of pure and mixed race, tribal affiliation and languages, who number about 133,000 of the total population (D.A.A., 1976:60).

In view of the diversity of all these aspects of the population, it is difficult to support the current usage of the concept of multicultural Australia. This appears to have originated as much from a politically expedient ideology, subsequently to become conventional wisdom, as from any rigorous analysis of what it really means.

The concept is fraught with several difficulties. The meaning of culture is not clear. If it is used in the popular sense of aesthetic pursuits and 'folk' traditions and customs then Australia might possibly warrant the label multicultural, but it is likely to generate little of educational significance apart from trivialised, 'folksy' elements in the curriculum. It also begs the difficult question as to whether there is a majority 'Australian' culture and, if so, what its distinguishing features are. If culture is used in a more technical sense analogous to the definition proposed above, then the implication is that Australian society comprises a number of distinctive social groups, each with its culture as a problem-solving design for living. This may be a more realistic model but it is still too simplistic and cannot be supported empirically. It is obvious that all social groups in Australia share many cultural elements in common, i.e., that despite sub-cultural variations particularly in cultural values and traditions, there is a common cultural substratum. This comprises such aspects as the political system and bureaucracy, social welfare provisions, an official language of commerce and general social transactions (English), common technological and material elements in housing, domestic appliances, tools, and other standardized artifacts. The more ideological side of culture — its values, traditions, beliefs, norms and attitudes — may be subject to sub-cultural variation. Following this line of argument Australian society could be called multi-sub-cultural, but this is a cumbersome neologism and still not very useful.

Two solutions to this conceptual problem have emerged in the overseas literature in recent years. Despres (1968) has suggested that countries like Australia and the United States might best be thought of as heterogeneous rather than fully culturally pluralist. A heterogeneous society is one composed of local 'cultural sections' at the town, neighbourhood and possibly state level. They are differentiated according to their local institutions, which structure activities and express cultural values within the context of local communities. Only when local cultural sections are linked to the national level by 'broker institutions' does a society merit the description of cultural pluralism. More recently, the concept of multi-ethnicity or poly-ethnicity has been wedded to the heterogeneous model in an upsurge of interest in ethnicity in the United States and Great Britain (Despres 1975; Glazer & Moynihan, 1975). The demands of this ethnic revitalization movement are twofold: (a) concern for equality (economic and political); (b) concern for the preservation and maintenance of ethnic identities and distinctive cultures (White, 1976). It may be no accident that this 'new pluralism' has been a feature of the sixties and seventies. Besides coinciding with the civil rights movement in the United States and the collapse of much of the 'old order' following the withdrawal from Vietnam, the ethnic revitalization movement has found receptive soil in the doctrine of equal opportunity of education that has been a feature of much educational ideology in Western nations since World War Two.

In essence, the official ideology of educational equal opportunity has legitimized the politicization of ethnic consciousness. It is thus no accident that the new movement finds expression through legislation in the United States in what has become known as the Ethnic Heritage Studies Act. Title IX of the Elementary and Secondary Education Act (June 23, 1972, 86 Stat. 346) states at Section 901:

> In recognition of the heterogeneous composition of the nation and of the fact that in a multi-ethnic society a greater understanding of the contributions of one's own heritage and those of one's fellow citizens can contribute to a more harmonious, patriotic, and committed populace, and in recognition of the principle that all persons in the educational institutions of the Nation should have an opportunity to learn about the differing and unique contributions to the national heritage made by each ethnic group, it is the purpose of this title to provide assistance designed to afford to students opportunities to learn about the nature of their own cultural heritage, and to study the contributions of the cultural heritages of the other ethnic groups of the Nation.

Despite the emphasis on cultural factors, White makes it clear that 'a close reading of the testimony at the hearings for this Bill suggests that scholars and those representing various interest groups expressed additional concerns [for] ... the preservation of cultural identity as an end in itself, as well as a means to attaining equality of achievement' (White, 1976:11-12).

The likelihood of a similar movement occurring generally in Australia is problematical. Cox (1976) has discussed the complex diversity of competing loyalties in ethnic groups, lack of uniformity in their social patterns, their high degree of internal socio-economic stratification, and similar factors which seem to suggest less than unanimity among ethnics about their preferences for retaining either their ethnicity or their cultures. 'Many immigrants simply do not wish to be identified as 'Immigrants" and as "aliens", and they try very hard to become Australian by de-emphasising their ethnic connections' (Cox, 1976:114). Ethnic groups are at times indistinguishable from other (non-ethnic) pressure or interest groups in the community.

However, this is only one side of the picture. Since the work of Barth (1969), ethnicity has been seen by most sociologists to have two sides. There is firstly the subjective state of 'belongingness' or self-identification with the ethnic in-group held by its members. Secondly there are the observable criteria such as language, traditions and customs, religious affiliation, which are externally observable evidence of the existence of the ethnic group. In other words, ethnic identity may be self-ascribed, ascribed by others, or both (Despres, 1975: 193). On these grounds we can usefully refer to Australia as multi-ethnic or poly-ethnic — the latter being preferred to make confusion with multi-cultural less likely and on grounds of strict, grammatical accuracy. The Australian population comprises a variety of ethnic groups, whose members may have self-ascribed ethnic identity or have it imposed on them by others, i.e., by 'native-born Australians'. The propensity for these to attach perjorative, stereotyped labels to 'foreigners' has been shown by Yarwood (1964:43). Beswick and Hills (1972) found 'a broad we-they identification in Australians with reference to immigrants. That is to say, there is a noticeable sense among "dinkum" Australians that they form a "we" group (an in-group), and that immigrants form a discriminable "they" group (an out-group)' (Kovacs & Cropley, 1975:64). In other words, ethnics are labelled as such willy-nilly.

The existence of this phenomenon of socially ascribed ethnicity, and consequent inferior status and discrimination in Australia, may precipitate the politicization of ethnicity that has occurred in the United States. Among members of the Greek (Papadopoulos u.d.) and Italian (Bertelli, u.d.) communities at least there is evidence of a movement concerned with gaining a more equable share of social and economic resources for immigrants. Among Aborigines there is also growing political consciousness, and such organizations as the National Aboriginal Congress and National Aboriginal Education Committee are striving for a fairer share of resources. Although it is still wedded to the concept of multi-culturalism, a recent publication of the Office of Community Relations (1976) makes clear that it is concerned to alleviate discrimination in such areas as law, education and social welfare.

## Problematics in the curriculum

If we concentrate on education, however, it is far from clear that current thinking about the education of children from ethnic backgrounds will in fact remove much of the discrimination they currently face. There are theoretical grounds for thinking that the majority group in a pluralist society, i.e. the 'Staatsvolk', is also an ethnic group (Connor, 1973:2-3). Following this line of argument 'dinkum' Australians might well constitute an ethnic group in competition for resources with other ethnic groups. As its members control the education system, they are well placed to ensure that children from other ethnic groups face educational hurdles that 'dinkum' children do not face to the same degree. There may be nothing malicious or really intentional about this type of educational discrimination. There does not need to be: the problematical aspects of the curriculum — the strategic arm of formal enculturation — works in its favour.

### (a) Problematics in the communication system

Most attention in migrant education has been given to English language deficits on the part of children from ethnic backgrounds. It is obviously an important aspect of the curriculum but has thrown undue attention on the purely linguistic modes of cultural communication, to the neglect of paralinguistic and extra-linguistic modes. A great deal of cultural information is signalled through them. At a fairly obvious level of analysis, teachers with ethnic children can say one thing linguistically, but can simultaneously negate the message by inappropriate paralinguistic or kinesic means. These modes are regularly used to signal information about the rules and contingencies of relationships. We can say one thing but demonstrate the opposite. The cultural message does not get through to the ethnic child, or does so in garbled form. One way of reducing this risk is to treat the English language as but one aspect of communicative competence (Hymes, 1971) in which as much importance is attached to knowing the social rules for using English as to the actual syntactic and phonological skills.

At a deeper level of analysis, the efficacy of any mode of communication depends on the recipient's ability to recognize it as a message system of significance to him- or herself. Huttar has shown in the case of Aboriginal children in schools in the Northern Territory, that a great deal of their learning is by looking and doing (Huttar, 1977), not by listening to teacher's purely verbal exposition or by carrying out individually conducted experimentation and inquiry-based, auto-didactic learning. A teacher's 'message' may be totally ignored. Similarly, other 'messages' in display material may not be seen as relevant, and may be also ignored.

The capacity to receive, decode and act upon messages is culturally determined in more subtle ways. We convey meaning through rituals, and these make up a great deal of what goes on in schools as Bernstein, Elvin and Peters suggest (1971). Some rituals are culturally universal in the sense that they can be understood by all, regardless of ethnic background. Other rituals have cultural content that makes them specific to the ethnic group using them. They may not be understood by others.

Other messages are signalled by the context of what is said or done, by what Bateson has termed 'meta-communicative frames' (1973:150 ff.). For instance, the curtain and stage of a play signal the message that what is performed should not be taken as reality. In a cross-cultural audience with children from a variety of ethnic backgrounds we face the possibility that this cultural convention may not be understood. The message in the performance is taken for reality when in fact it is meant to be fiction. The same phenomenon is known to affect very young, kindergarten children watching the Six Million Dollar Man's antics on television. The 'message' of the context — 'this is only a play' — is not received, with the consequence that little children attempt to replicate the leaps and other superhuman feats of the actor, and get injured. Other signals, usually of a kinesic or paralinguistic kind, convey messages about the mood of the person communicating information. This again is a cultural phenomena. A purely linguistic message may fail to get through to an ethnic child, if the other signals he receives which set the context of the message convey a mood that is completely at odds with what is said. For instance, a teacher who says 'I like you in my class', but conveys something totally different by rigid body posture, facial expression and other kinesic signals is unlikely to be believed.

A further problematic aspect of the curricular communication system is due to the iconic nature of much of our communication (Bateson, 1973:382ff.). A great deal of information is signalled in an incomplete fashion on a part for whole basis, which creates the phenomenon of redundancy. 'The message material is said to contain "redundancy" if, when the sequence is received with some items missing, the receiver is able to guess at the missing items with better than random success' (ibid., p.389). But ability to fill in the gaps in a message is a matter of being familiar with the overall pattern of signals from past experience. This, of course, is a matter of cultural training. A child from a different cultural background will not be able to supply the missing information created by redundancy, and consequently may miss the full significance of the message. Other parts of it may constitute 'noise' which those who know the communication style habitually filter out as irrelevant. At the most extreme we can think of a speaker's 'ums', 'ahs', and other interjections in the category of 'noise'. A listener who does not know what these are will attach undue significance to such utterances as if they were important parts of the message.

Once a message and information has been received in the form of cognitions, it has still to be processed into a person's 'cognitive map; (Hill, 1963:114-128). That is, according to this learning theory, we learn by adding cognitions to our cognitive maps. This is only achieved with certainty if there is some regularity about the world, i.e., we can expect the world to be organized in a certain way and not others. Our expectations based on past experience are 'field expectancies' which are cognitions about the way the world is laid out or arranged. These make up a great deal of our knowledge. The phenomenon of 'culture shock' or 'reality shock' (Holzner, 1972:56) occurs when previously learned cognitions do not accord with new experience. It can affect both children and adults from ethnic backgrounds, particularly if their cultures are markedly different from the Australian as would be the case, for instance, when a Vietnamese child from a rural village is brought over to a hostel in suburban Melbourne, and attends the local school.

A related feature of cultural differences in processing information is suggested by recent research by Witkin and associates (Witkin, 1977). This has been taken up by Ramirez and Castaneda (1974) who have examined different learning styles and their relationship to cultural diversity. They suggest that there is a connection between one's culture and one's intellectual and affective behaviours, which follows the broad distinction between field dependency and field independency established by the Witkin team. The former relates to people whose cognitive styles are strongly dominated by the prevailing field, i.e., by the overall organization of phenomena and by external cues. Such people are sensitive to social cues, interested in what others do and say, are drawn to people, physically employ low social distance, and are interested in selective social aspects of their surroundings. Field independent people are those whose cognitive styles enable them to experience phenomena more or less independently of or separately from the surrounding field. Such persons are held to be more impersonal, not sensitive to social undercurrents, cool, distant, and unaware of their social stimulus value.

A final feature of children's learning that is also culturally determined must be taken into account. This can be termed deutero-learning, following the earlier work of Bateson (1958), subsequently elaborated in his later work (1973). As adapted by the writer (1975) deutero-learning can be thought of as learned styles of processing information and dealing with learning contexts. Evidence for cross-cultural differences in deutero-learning was provided by the

Jewish boys which were the subject of this piece of research. They employed traditional Orthodox rote learning techniques accompanied by sotto voce verbalization and body swaying, in learning contexts which demanded reasoning ability and the application of learned information to new, problem-solving situations. It was suggested that inappropriate deutero-learning techniques would set up considerable dissonance in the boys because they were operating at a largely unconscious level and were thus not amenable to conscious attempts to reduce dissonance. Other examples can be found in aboriginal children who learn a great deal of kinship information aurally by rote repetition. Learning new psychomotor skills might also be affected by deutero-learning. Rosalie Wax (1967) has described how a Mayan girl learned how to operate a weaving machine being demonstrated to her class. She watched intently, but would not 'have a go' for fear of making a mistake and being shamed in front of her peers. Instead of learning by trial and error, on the assumption that practice makes perfect (an accepted Western way of learning a motor skill) she apparently 'imprinted' the correct sequence in her mind, and was able to perform the correct operations faultlessly when she finally made an attempt at weaving on her own.

(b) Problematics in the power system

It should be apparent from the above discussion that the teacher is unconsciously able to exert considerable power over the learning of ethnic children if he or she is ignorant of the various factors that make it a problematical enterprise in an intercultural classroom. In this section of the paper, however, we are more concerned with the exercise of power along the lines suggested by Etsioni, at the three levels of the curriculum.

At the level of curriculum selection there is the risk that those who devise the curriculum will stress that knowledge that favours their own socio-cultural milieu, and will neglect or devalue that of the ethnic groups present in the school. This writer has shown how a special category of knowledge (super-empirical knowledge) formed an important part of the religious curriculum of the Orthodox Jewish Ethnic Day School he studied (Bullivant, 1975). Although the example is exceptional, it should not be dismissed outright. The growing number of Turkish, Arabic and other believers in Islam in our poly-ethnic society raises similar problems. To cater for them should we include selections from the Koran in any curriculum which includes a moral education component ? One result of doing so might be to confront a child with two views of social or natural reality. One stems from fundamentalist Biblical or Koranic scripture, the other from Western science. In the case of the writer's research, such a situation invaded even seemingly 'safe' geography lessons when the age of rocks and evolution of the world were discussed. To insist upon the absolute rightness of Western scientific thought would have been an exercise in cultural power.

Unfortunately the teacher is not the sole arbiter of what to select for the curriculum. He is subject to supra-systemic pressures from organizations such as his subject association and teachers' unions. These have vested interests to protect and subject empires to defend, and may take little note, if any, of the wishes and needs of the ethnic pupils in the school. We thus have anomalous situations where children fluent in Italian, Greek, or Serbo-Croation are forced to learn French because this subject is an established one in the curriculum.

Even more bizarre examples can be found, such as the aboriginal girls from fringe settlements in the Northern Territory learning the mysterious art of making a pavlova because it was prescribed on the domestic science syllabus set by the educational administration.

At the organizational level opportunities exist for controlling ethnics' access to knowledge through the use of remunerative power. Again, this may be exercised deliberately or unconsciously. Research into New South Wales and Victorian schools by the Committee for the Inquiry into Schools of High Migrant Density (1976) reveals very great deficits in specialist staffing, teaching materials, accommodation for migrant classes and other resources. Yet there is also evidence to suggest that these can usually be found where the teaching of native-born 'Australian' children is concerned. It is difficult to escape the conclusion that either the problems of ethnics are not recognized, or that differential allocation of resources is being made in ways that work against their interests.

The matter goes deeper than this however. The level of curriculum organization involves making value judgements about the sequencing and grouping of learning for the child. Decisions have to be made, for instance, about the time to be allocated to any unit of work. For those who find it difficult to cope with core knowledge and experiences in the curriculum Bloom suggests (1971:17) that a mastery learning approach should be used. The key variable to help the under-achiever is his willingness to persist and the teacher's willingness to spend more time encouraging progress through the appropriateness and variety of instruction. Time, however, is a scarce resource and it may not be possible to allocate more time to slow-learning ethnics. The matter also involves the grouping of classes. It may be necessary to retard the promotion of ethnic children from grade to grade to ensure that mastery learning is completed in one grade before proceeding to the next.

Types of grouping should also involve consideration of children's personalities. This would appear to be even more desirable in view of the known connection between culture and personality (e.g. Honigmann, 1967). Eysenck has suggested (1977) that 'setting' might be used for special assignments and novel teaching practices more suitable for extraverts, who could be allowed to work on their own by discovery methods, while the introverts in the class would continue to be taught by more traditional, didactic methods. But such organizational decisions again involve re-allocating classroom and other resources, and may not be seen as warranted, even if the school administrators are able to recognize culturally determined personality differences among ethnic children.

At the level of curriculum implementation the control of power is more personalized as it involves face-to-face transactions between the teacher and children in his or her care. This offers opportunities for the teacher to exercise influence over ethnic children by restricting their access to ways in which they can express and re-affirm their ethnic heritage. This is all too easy. For instance, the name of a Greek child is intimately connected with family honour and the kinship group. The name is invested with considerable symbolic significance, and an uncaring or lazy teacher who cannot or will not pronounce a Greek child's name correctly is exercising normative power by reducing the child's self-esteem and feelings of ethnic pride. A similar effect can be achieved by forbidding the child from using his native language which is known to be intimately connected with culture and self-identity (Holzner, 1972:33 ff.).

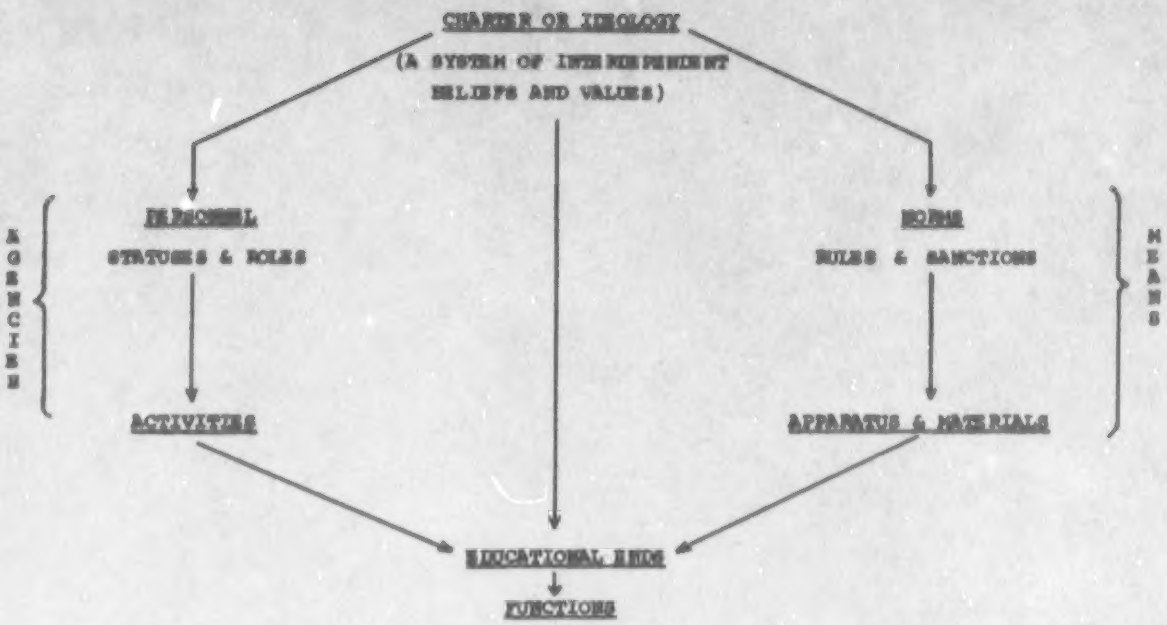
Control over power is very obviously involved in keeping discipline in a school, but in the poly-ethnic classrooms it is not as straightforward as it might appear. Teacher's control involves considerable use of normative power in making rewards and imposing sanctions. However, a teacher's rewarding behaviour may not be recognized as such by ethnic children. In turn, the teacher may not recognize their positive feedback if it is given in a communication style with which he/she is unfamiliar. The negotiation and bargaining, described by Geer (1968) as an essential part of classroom control, are workable only if teacher and pupils share the same frame of reference or assumptions about normative power and discipline, and know the same rules of the bargaining game. In the poly-ethnic classroom this again is problematical.

## CONCLUSION

The curriculum has been discussed in relation to the poly-ethnic society as an essentially problematical enterprise, and a number of examples have been given to illustrate this general thesis. In looking to the future in Australian education the problems discussed suggest that much more research might be invested into looking at the cultural factors involved in at least two aspects of the curriculum: as communication system and as power system. The poly-ethnic composition of Australian society is unlikely to diminish, and the challenges it poses should be faced squarely through the types of analysis provided by anthropology of education, rather than dodged. Ultimately all our children would stand to gain from an understanding of the curriculum problematics that stem from the type of socio-cultural context in which they are growing up.

FIGURE 1

EDUCATION AS AN INSTITUTION



CHARTER OR IDEOLOGY
(A SYSTEM OF INTERDEPENDENT BELIEFS AND VALUES)
PERSONNEL
STATUSES & ROLES
ACTIVITIES
NORMS
RULES & SANCTIONS
APPARATUS & MATERIALS
AGENCIES
MEANS
EDUCATIONAL ENDS
FUNCTIONS

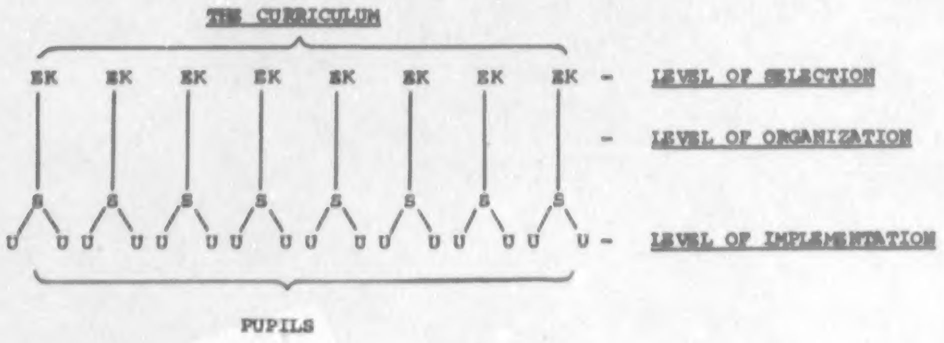
(Bullivant, 1977 after Malinowski)

---

FIGURE 2

A CULTURAL MODEL OF THE CURRICULUM

A SELECTION FROM A SOCIO-CULTURAL GROUP'S STOCK OF VALUED TRADITIONAL AND CURRENT PUBLIC KNOWLEDGE (K) AND EXPERIENCES (E) USUALLY PURPOSEFULLY ORGANIZED BY SUCH EDUCATIONAL AGENCIES AS SCHOOLS, AND MADE AVAILABLE TO PUPILS SEQUENTIALLY THROUGH SYLLABUSES (S) AND UNITS (U)



THE CURRICULUM
EK EK EK EK EK EK EK EK
- LEVEL OF SELECTION
- LEVEL OF ORGANIZATION
S S S S S S S S
U U U U U U U U U U U U U U U U
- LEVEL OF IMPLEMENTATION
PUPILS

## References

Banks, J.A. 1975 *Teaching Strategies for Ethnic Studies*. Boston: Allyn and Bacon.

Banks, A. *et al*. 1976 The NCSS Task Force on Ethnic Studies Curriculum Guidelines, 'Curriculum Guidelines for Multiethnic Education', *Social Education*, October, Washington: N.C.S.S.

Barth, F. 1969 *Ethnic groups and boundaries*. Boston: Little, Brown.

Bateson, G. 1958 *Naven* (Second edition). Stanford, Calif.: Stanford University Press.

Bateson, G. 1973 *Steps to an Ecology of Mind*. St.Albans, Herts.: Granada (Paladin Books).

Bernstein, B. 1971 'On the Classification and Framing of Educational Knowledge'. In M.F.D. Young (ed.) *Knowledge and Control*. London: Collier-Macmillan.

Bernstein, B., Elvin, H.L. and Peters, R.S. 1971 'Ritual in Education'. Pp.160-165 in The School and Society Course Team (eds.), *School and Society. A sociological reader*. London: Routledge & Kegan Paul, Open University Press.

Bertelli, L. u.d. 'Social Organization and Ethnic Power: an Italian Perspective'. Pp.35-38 in D. Cox (ed.) *Ethnic Rights Power and Participation*. Melbourne: Clearing House on Migration Issues, Ecumenical Migration Centre, and Centre for Urban Research and Action.

Beswick, D.G. & Hills, M.D. 1972 'A survey of ethnocentrism in Australia'. *Australian Journal of Psychology*, 24, pp.153-164.

Bloom, B.S. 1971 'Mastery learning and its implications for curriculum development'. In E.W. Eisner (ed.) *Confronting Curriculum Reform.* Boston: Little, Brown.

Bourdieu, P. 1971 'Systems of education and systems of thought'. In M.F.D. Young (ed.) *Knowledge and Control*. London: Collier-Macmillan.

Bronfenbrenner, U. 1974 *Two worlds of childhood (USA and USSR)*. Harmondsworth: Penguin.

Bullivant, B.M. 1972 'The cultural reality of curriculum development'. *Education News*. 13(9), June, pp.14-17.

Bullivant, B.M. 1973a 'Is there a hidden curriculum in curriculum development?', *Twentieth Century*, 27(3), Autumn, pp.239-253.

Bullivant, B.M. 1973b *Educating the Immigrant Child: Concepts and Cases*. Sydney: Angus & Robertson.

Bullivant, B.M. 1975 Values and Traditions in an Orthodox Jewish Day School. A Study of Enculturation Dissonance. Unpublished Doctoral Thesis. Melbourne University.

Bullivant, B.M. 1978a 'The schooling of ethnic children'. In R.King & J.McArthur (eds.) Educational Administration in Australia, (working title). Sydney: McGraw-Hill (forthcoming).

Bullivant, B.M. 1978b Race and Curriculum: an Australasian Perspective. Melbourne: Australia International Press, (forthcoming).

Committee for the Inquiry into Schools of High Migrant Density 1976 *Inquiry into Schools of High Migrant Density: 1974 study based on Schools selected in New South Wales and Victoria*. Canberra: Australian Government Publishing Service.

Connor, W. 1973 'The politics of ethnonationalism'. *Journal of International Affairs*, 27(1) pp.1-21.

Cox, D.R. 1976 'Pluralism in Australia'. *The Australian and New Zealand Journal of Sociology*, 12(2), June, pp.112-117.

Dahlke, H.O. 1958 *Values in Culture and Classroom. A Study in the Sociology of the School*. New York: Harper & Brothers.

Department of Aboriginal Affairs 1976 *Annual Report 1975-76*. Canberra: Australian Government Publishing Service.

Despres, L.A. 1968 'Anthropological Theory, Cultural Pluralism, and the Study of Complex Societies'. *Current Anthropology*, 9(1): pp.3-26.

Despres, L.A. (ed.) 1975 *Ethnicity and Resource Competition in Plural Societies*. The Hague: Mouton.

Etzioni, A. 1969 'A basis for comparative analysis of complex organizations'. In A. Etzioni (ed.) *A Sociological Reader in Complex Organizations*. Second Edition. New York: Holt, Rinehart and Winston.

Eysenck, H.J. 1977 'The Development of Personality and its Relation to Learning'. Fink Memorial Lecture. Melbourne: University of Melbourne.

Frake, C.O. 1973 'The Diagnosis of Disease among the Subanun of Mindanao'. Ch.7 in N. Keddie (ed.) Tinker, tailor ... the Myth of Cultural Deprivation. Harmondsworth: Penguin.

Geer, B. 1968 'Teaching'. Pp.560-5 in D.L. Sills (ed.) International Encyclopedia of the Social Sciences. New York: Free Press.

Gladwin, T. 1973 'Culture and Logical Process'. Ch.6 in N. Keddie (ed.) Tinker, tailor ... the Myth of Cultural Deprivation. Harmondsworth: Penguin.

Glazer, N. & Moynihan, D.P. 1963 Beyond the Melting Pot. Cambridge, Mass.: M.I.T. Press.

Hill, W.F. 1963 Learning: a survey of psychological interpretations. San Francisco: Chandler.

Holzner, B. 1972 Reality Construction in Society. (Revised Edition). Cambridge, Mass.: Schenkman.

Honigmann, J.J. 1967 Personality in Culture. New York: Harper & Row.

Horton, R. 1971 'African Traditional Thought and Western Science'. Ch.8 in M.F.D. Young (ed.) Knowledge and Control. London: Collier-Macmillan.

Hostetler, J.A. 1970 'Socialization and Adaptations to Public Schooling: The Hutterian Brethren and the Old Order Amish'. The Sociological Quarterly, 11(2): pp.194-205.

Hostetler, J.A. & Huntingdon, G.E. 1971 Children in Amish Society: Socialization and Community Education. New York: Holt, Rinehart and Winston.

Huttar, G.L. 1977 'World Views, Intelligence and Cross-cultural Communication'. Ethnic Studies, 1(3).

Hymes, D. 1971 'On Linguistic Theory, Communicative Competence, and the Education of Disadvantaged Children'. In M.L. Wax, S. Diamond & F.O. Gearing (eds) Anthropological Perspectives on Education. New York: Basic Books.

Keddie, N. 1971 'Classroom Knowledge'. Pp.133-160 in Young, M.F.D. (ed.), Knowledge and Control: New Directions for the Sociology of Education. London: Collier-Macmillan

Keddie, N. (ed.) 1973 Tinker, Tailor ... the Myth of the Deprived Child. Harmondsworth: Penguin.

Kovacs, M.L. & Cropley, A.J. 1975 Immigrants and Society: Alienation and Assimilation. Sydney: McGraw-Hill.

Lawton, D. 1975 Class, Culture and the Curriculum. London: Routledge & Kegan Paul.

Leach, E. 1976 Culture and Communication. The logic by which symbols are connected. Cambridge: Cambridge University Press.

Malinowski, B. 1944 A Scientific Theory of Culture. Chapel Hill: University of North Carolina Press.

Office of Community Relations 1976 First Annual Report 1976 of the Commissioner for Community Relations. Canberra: Australian Government Publishing Service.

Papadopoulos, G. u.d. 'Social Organization and Ethnic Power: a Greek Perspective'. Pp.39-46 in D. Cox (ed.) Ethnic Rights Power and Participation. Melbourne: Clearing House on Migration Issues, Ecumenical Migration Centre, and Centre for Urban Research and Action.

Ramirez, M. & Castaneda, A. 1974 Cultural Democracy, Bicognitive Development and Education. New York: Academic Press.

Reynolds, J. and Skilbeck, M. 1976 Culture and the Classroom. Melbourne: Macmillan.

Smith, B.O., Stanley, W.O. and Shores, J.H. 1957 Fundamentals of Curriculum Development. New York: Harcourt, Brace and World.

Taba, H. 1962 Curriculum Development, Theory and Practice. New York: Harcourt, Brace and World.

Theodorson, G.A. & Theodorson, A.G. 1970 A Modern Dictionary of Sociology. London: Methuen.

Wax, R.H. & Thomas, R.K. 1967 'American Indians and White People'. Ch.14 in E.T. Keach et al. (eds) Education and Social Crisis. New York: Wiley.

Wheeler, D.K. 1967 Curriculum Process. London: University of London Press.

White, N.R. 1976 Cultural Pluralism and Schooling. Ph.D. Thesis, Graduate School of Syracuse University. Unpublished.

Witkin, H.A. et al. 1977 'Field-Dependent and Field-Independent Cognitive Styles and their Educational Implications'. Review of Educational Research. 47(1), Winter.

Yarwood, A.T. 1964 Asian Migration to Australia. Melbourne: Melbourne University Press.

Young, M.F.D. (ed.) 1971 Knowledge and Control. London: Collier-Macmillan.

# The Validity of Alternative Methods of Scaling School Assessments at the HSC Level for the Colleges and High Schools of the Australian Capital Territory

J.P. Keeves, B. McBryde and L.A. Bennett
Australian Council for Educational Research

Three successive reports, Currie (1967), Campbell (1973) and the ACER Report Thomson (1974) on education in the Australian Capital Territory have advocated change in the system of public examinations at the terminal secondary school stage. With an estimated 54 percent of the age cohort remaining at school to the Year 12 level and with the introduction, in 1976, of Secondary Colleges for students at Years 11 and 12, it was highly desirable that new approaches to the curriculum should be developed and innovatory courses should be injected into the school system. Nevertheless, it was essential that some measure of performance during these final two years of secondary schooling should be available to assist with selection for tertiary education.

The approach adopted by the ACT Schools Accrediting Agency which was set up by the ACT Schools Authority was to provide a profile of information on student performance in a Secondary College Record and, in addition, to provide Supplementary Information for tertiary selection. It is not the purpose of this paper to describe in full detail the procedures employed for the accreditation of courses, or to consider the range of information to be provided on the Secondary College Record and as Supplementary Information. One critical item - the Tertiary Entry Score - is, however, of considerable interest because there are important theoretical and practical problems associated with its calculation. A Tertiary Entry Score was initially introduced to give portability outside the ACT to the transcript provided for each student by the Agency. However, the Australian National University and the Canberra College of Advanced Education have also decided to make use of this score in the selection of students to enter first year tertiary courses in 1978. Consequently, it has become increasingly important that the procedures for calculating the Tertiary Entry Score should not only be valid, but should also be accepted by students, teachers, parents and by the admissions committees of all tertiary institutions which admit students from schools in the ACT.

Within the ACT there are approximately 2100 students in 14 institutions, which are currently conducting Year 12 courses, and a majority of these students at this level are in the five Secondary Colleges. Each institution has developed its own program of courses, which have been accredited by the ACT Schools Accrediting Agency, and it is important to note that a wide variety of courses have been offered during 1976 and 1977, which have catered for students of differing levels of ability and scholastic aptitude. The Currie Report (1967) and the Campbell Report (1973) both recommended the policy of assessment within the secondary school which was adopted by the ACT Schools Authority in preference to the use or the establishment of a system of public examinations. Thus, the major problem of measurement that confronted the agency was the calculation of a Tertiary Entry Score from the teachers' assessments of student performance on a wide range of courses in the 14 colleges and high schools.

The evidence presented by McGaw (1975) and McGaw, Warry and McBryde (1975) from studies undertaken in Queensland indicated that it was possible to rescale teacher assessments using the Australian Scholastic Aptitude Test to yield measures of performance of strong validity and reliability. However, the situation in the ACT differed from that which existed in Queensland, in so far as the ACT schools and colleges had a wider range of courses, which were being taken in a considerably smaller number of schools. As a consequence, it was decided to abandon the two stage rescaling procedure which was being employed in Queensland, whereby after rescaling of subject assessments an aggregate score was first calculated within schools, and subsequently these aggregate scores were rescaled for differences between schools. It was planned to employ a simpler one stage procedures in which rescaling both between schools and between courses within schools would be carried out in a single step using performance on the Australian Scholastic Aptitude Test. This procedure is illustrated by Figure 1. which records the operations being employed for the rescaling of school assessments on each course offered.

As a first step each student requiring a Tertiary Entry Score, must sit for the Australian Scholastic Aptitude Test consisting of two papers each of 1½ hours length with a total of 100 items. In addition, each secondary college or high school provides an assessment of the performance of each student on each course taken, on a scale with a mean of 65 and a standard deviation of 15. After the rescaling of the scores for each course within each college or high school an aggregate of the course scores is calculated using performance on three major courses, with a maximum of 100 marks per course, and one minor course, with a maximum of 60 marks, or their equivalent, to yield a total score with a maximum value of 360. The calculation is carried out in such a way as to select from the course scores obtained by each student those scores that maximize his total score.

In the rescaling of school assessments there are a number of important issues which must be examined to ensure that the procedures employed are valid. More importantly, since any rescaling procedure is subject to a degree of uncertainty, it is necessary to consider the procedures being used in comparison to alternative methods.

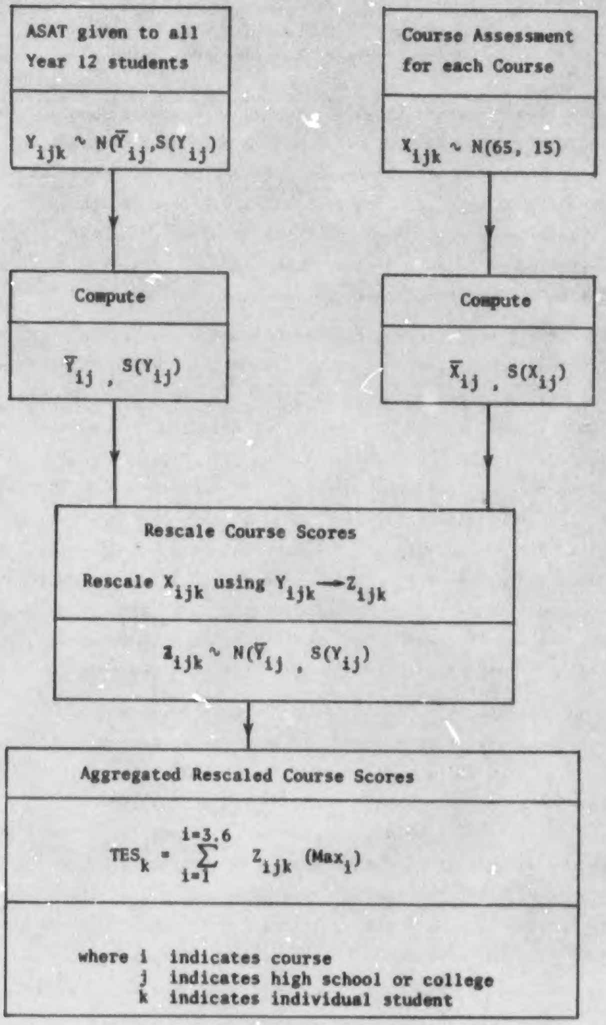

ASAT given to all Year 12 students
$Y_{ijk} \sim N(\overline{Y}_{ij}, S(Y_{ij}))$
Course Assessment for each Course
$X_{ijk} \sim N(65, 15)$
Compute
$\overline{Y}_{ij}\ ,\ S(Y_{ij})$
Compute
$\overline{X}_{ij}\ ,\ S(X_{ij})$
Rescale Course Scores
Rescale $X_{ijk}$ using $Y_{ijk} \rightarrow Z_{ijk}$
$Z_{ijk} \sim N(\overline{Y}_{ij}\ ,\ S(Y_{ij}))$
Aggregated Rescaled Course Scores
$TES_k = \sum_{i=1}^{i=3.6} Z_{ijk}\ (Max_i)$
where i indicates course
j indicates high school or college
k indicates individual student

Figure 1. Procedure for Rescaling School Course Scores

## Issues of Validity

In the above procedure for rescaling school assessments, a large number of disparate measures, obtained on a wide range of courses, must be brought to a common scale prior to adding to yield a Tertiary Entry Score. The process of rescaling the measures for different courses onto a common scale is frequently questioned. However, it is important to note that unless the measures are on a common scale, the operation of addition to obtain a total score is scarcely valid. Clearly, the decision to use a total score requires that the different measures must be tied to a common scale. It should be noted, however, that the ACT Schools Accrediting Agency has constantly questioned the sole use of a total Tertiary Entry Score for selection to tertiary institutions on the grounds that all course scores are being brought to a common scale, although different developed abilities and aptitudes may be required for the study of different courses at the tertiary level. It is with this in mind that the Agency has planned to present a profile of information and thus to invite tertiary institutions to use the Tertiary Entry Score, and all other information provided, in appropriate ways.

The Agency has found it necessary to acknowledge that the procedure of rescaling all courses to a common scale in order to form a total score assumes that underlying academic performance is a common factor of scholastic aptitude or developed ability that is equally strongly related to performance in all subject areas. Moreover, it must be assumed that by employing the Australian Scholastic Aptitude Test for the purposes of rescaling, this common factor is adequately measured by this test.

The evidence available would appear to indicate that there is a single general factor of scholastic aptitude built into the ASAT tests. The between student reliabilities of these tests are, in general, not less than 0.90 (Morgan, 1977). Moreover, a study, recently completed by Bell (1977), has provided further strong support for this contention. The face validity of the more recent ASAT tests has been high, but earlier versions of the TEEP and ASAT tests had spatial and aesthetic components that were of more doubtful strength. Nevertheless, the validity of the test as a moderating variable for the purposes of rescaling has been questioned by Brennan (1976) and others, because of relatively low correlations with many matriculation subjects. The low correlations which may indicate that there is not a single general factor underlying academic achievement in all subject areas, must be considered with respect to how the moderating or reference test is to be used in the process of rescaling. It is clearly possible, as has been pointed out,

> the use of moderating tests masks differences in mean performance of groups caused by factors other than the "quality" or "ability" of the group as measured by the tests. These factors may include the quality of instruction received, socio-economic factors prevailing or psychological traits (other than ability) which influence performance (Cook and Cooney, 1976:7).

These possibilities are too strong to be lightly cast aside and ignored.

## Procedures of Rescaling

Cooney (1975) following Howard (1958) has identified two requirements that rescaling procedures should satisfy. First, the students whose marks are being rescaled should not be disadvantaged by the marking patterns of their assessors in the colleges. This implies that the spread of marks, as indicated by the standard deviation, and the shape of the distribution of course assessments provided by the high school or college, may need to be adjusted in the light of further evidence of a more objective kind. Secondly, the student should not be disadvantaged by characteristics of the candidature with whom he competes. This implies that the mean scores of the distributions of the course assessments may need to be adjusted in the rescaling operation according to the level of developed ability or scholastic aptitude of the student groups.

It is important to note, as Cooney (1975) has pointed out, that if the course assessments are obtained by the averaging and adding of a large number of component scores, then provided the underlying characteristic of performance is normally distributed in the student group, the course assessment scores might, by the Central Limit Theorem, be expected to be normally distributed.

Three general procedures are currently under consideration by Australian examining authorities for the rescaling of marks: (i) an additive procedure, (ii) a linear transformation procedure, and (iii) the equipercentile procedure. While the rescaling operations employed in some Australian States do not follow exactly the methods outlined below they are variants of one of these three procedures.

1. The Additive Method for Rescaling In this procedure the assessment scores for different courses are initially scaled so that they all have the same mean and variance, under the assumption that the spread of developed ability involved (as indicated by the standard deviation of scores) is the same for each subject. In a second step, a component is added or subtracted to allow for the relative standing of the group of students with respect to some reference population. Thus a more able group of students taking a more difficult subject will all receive an additional score and a less able group of students will all have their scores reduced in accordance with the standing of the group as assessed by a common reference test or in other ways. It should be noted that this procedure does not change the spread or shape of the course assessment score distributions, and for this reason was, in general, rejected as being inappropriate for use in the ACT, where it could not be assumed that the spread and shape of the score distributions were the same across all courses and all high schools or colleges.[1]

In mathematical terms this method of rescaling may be recorded:

$$Z_{ijk} = X_{ijk} + D_{ij}$$

where $Z_{ijk}$ is the moderated course assessment score for student k in course j in school i,

$X_{ijk}$ is the original course assessment score for student k in course j in school i,

and $D_{ij}$ is the deviation for course j in school i as assessed by some common reference test or in other ways.

---

[1] A variant of this procedure is being used for the rescaling of school assessment scores with very small groups for whom the calculated standard deviations of scores are likely to be unreliable.

194

2. The Linear Transformation Method of Rescaling In this procedure the course assessment scores are rescaled using a reference test. The linear rescaling equation is based on the equality of the standard scores for the two distributions and the shape of the distribution of the course assessment scores is left unchanged. It is a basic premise of this method that in making course assessments teachers can make valid and reliable judgements concerning the relative performance of students within the courses in a school both with respect to the skewness properties and shape of the distribution of scores. It is also assumed that the teachers are unable to make sound estimates of either the relative standing of the group, or of the spread of performance within the group and that both of these characteristics must be moderated by means of some common reference test.

The procedure of linear rescaling may be stated in mathematical terms as follows:

$$Z_{ijk} = \overline{Y}_{ij} + \frac{S(Y_{ij})}{S(X_{ij})} (X_{ij} - \overline{X}_{ij})$$

where $Z_{ijk}$ is the moderated course assessment score for k student in course j in school i,

$X_{ijk}$ is the original course assessment score for student k in course j in school i with a school and course mean of $\overline{X}_{ij}$ and standard deviation of $S(X_{ij})$.

$\overline{Y}_{ij}$ is the mean of the reference test with respect to course j in school i for reference test scores $Y_{ijk}$ which have a standard deviation of $S(Y_{ij})$.

It should be noted that in this procedure the correlation between the course assessments and performance on the reference test is arbitrarily set at 1 for all courses and all schools, although it must be acknowledged that in practice these correlations deviate significantly from unity and differ markedly between courses and between schools. The effect of the procedure is to adjust the mean and standard deviation of the course assessment scores so that they coincide with those of the distribution of scores on the reference test. If the distribution of both the course assessment scores and the reference test scores are normal this method of rescaling does not differ from the equipercentile method of rescaling described in the next section.

3. The Equipercentile Method of Rescaling In this procedure where x and y are scores on course assessments (X) and on the reference test (Y) and F(x), G(y) their respective cumulative relative frequencies, then the equipercentile method of equating test scores states that x and y are equivalent, if F(x) = G(y). This transformation is an equal area or probability transformation and takes into account differences in both shape and spread of the two frequency distributions. Traditionally in equating test scores a graphic form of this procedure has been in common use. The techniques employed have been explained by Flanagan (1951). However, a graphical procedure is slow and tedious to work with. More recently Lindsay and Prichard (1971) have suggested an equivalent analytic procedure for use with a computer. It is important to note, that like the graphical approach, the analytic procedure is likely to be inappropriate when the number of cases is small and when the test score distribution is very seriously skewed. Nevertheless, where skewing is only moderate this procedure takes into consideration the shape of the score distribution of course assessments and the shape of the score distribution for the reference test of scholastic aptitude.

In considering procedures to be employed within the ACT for the rescaling of course assessments using as a reference test performance on the Australian Scholastic Aptitude Test, it was apparent that, while the Additive Method had little recommending its use, there was no clear evidence to assist in making a choice between the Linear Transformation Method and the Equipercentile Method. The fact that in Queensland the Linear Transformation Method was being used in a manner considered to be satisfactory (McGaw, 1975) could not be taken to imply that the Equipercentile Method would not be superior, since it would take into consideration both the shape of the distribution of scores associated with school assessments and those of the reference test of scholastic aptitude in a way that the Linear Transformation Method would not. However, it was recognized that where both distributions were approximately normal there would be little to choose between the two methods.

If an investigation were to be undertaken to examine the differences between these two rescaling procedures it would be necessary to incorporate into the study an appropriate criterion measure. A suitable criterion lay in the actual performance of the students on the Higher School Certificate Examination, but it was necessary to as whether the product moment correlation between students' scores on HSC subjects and the corresponding rescaled school assessments would provide the information that was sought to compare the effectiveness of the procedures being employed.

## The Choice of a Correlation Coefficient

Most studies that have been undertaken into the characteristics of the Australian Scholastic Aptitude Test both as a moderating variable for the rescaling of school assessments and as a predictor of concurrent and subsequent performance have reported results in terms of a between student correlation coefficient. Only the studies by McGaw (1975) and McGaw, Warry and McBryde (1975) have acknowledged that other correlation coefficients could provide useful information, and even those studies have not explored the full range of indices available.

Many years have now passed since Robinson (1950) first drew attention to the fallacies of examining data using between group correlations when between individual correlations were appropriate. Likewise, there is a corresponding fallacy in using between individual correlations when between group correlations are required to describe the situation. In the rescaling of school assessment marks and teachers' estimates of student performance, the units under consideration are both the individual student and the course within the school. Consequently, there are three distinct correlation coefficients which are involved in the description of an ecological situation. In mathematical terms these three correlations are described as follows:

The Between Student Correlation ($R_{zt}$) is the simple product moment correlation between the rescaled school assessment scores (Z) and the Higher School Certificate scores (T) for all members (N) of the total group, without reference (to courses or) to schools.

The Between School Correlation ($ER_{zt}$) is the weighted product moment correlation between the M pairs of mean scores on the rescaled school assessment scores (Z) and the Higher School Certificate scores (T) which describe the groups taking each course within each school.

The Between Student Within School Correlation ($WR_{zt}$) is the weighted average of the between student product moment correlations calculated between the rescaled school assessment scores (Z) and the Higher School Certificate Scores (T) within each of the M schools.

The relationship between the three correlations may be stated as follows (Hannan, 1971: 43).

$$ER_{zt} = \frac{R_{zt}}{E_{zr}\,E_{tr}} - \frac{WR_{zt}\sqrt{1 - E_{zr}^2}\sqrt{1 - E_{tr}^2}}{E_{zr}\,E_{tr}}$$

Where $E_{zr}$ and $E_{tr}$ are the two correlation ratios which indicate the clustering of the two measures, school assessment scores (Z) and Higher School Certificate scores (T) within the schools and courses.

Thus the between school correlation is the weighted difference between the between student correlation and the weighted average of the M between student within school correlations. The extent to which the between school correlation differs from the between student correlation is clearly related to the clustering of student performance within schools.

In order to assess the effectiveness of any rescaling procedure it would appear to be necessary to examine all three product moment correlations, between student, between school and between student within school, for not only the rescaled course assessment scores but also the initial unscaled course assessment scores, the scores on the moderating or reference test, and the scores on the criterion measure of the Higher School Certificate. Moreover, to explain the differences observed it would be necessary to consider the correlation ratios for these variables since they are associated with the degree of clustering within schools with respect to these variables. Before discussing these relationships in greater detail it is essential to consider the data which were available for examination and analysis.

## The Data

In 1975, as part of a preliminary study to investigate the characteristics of the Australian Scholastic Aptitude Test this test was administered to all students at the Year 12 level in the ACT. From this study data were also available for student performance on this reference test. These data could be linked to student performance on the Higher School Certificate Examination (1975) and to teacher assessment of school performance on Higher School Certificate courses. The measures of teacher assessment were made in terms of teacher expectation of performance on the HSC Examination and not by the aggregation and averaging of marks obtained in course work during the ultimate and pen-ultimate years of schooling.[2] Thus the teacher assessment scores had a common frame of reference across all schools. Moreover, in the New South Wales school system in 1975, it was possible, in general, to study courses, at three levels, and the Higher School Certificate Examination marks were brought to a common scale across the three levels. Consequently, teachers were accustomed to measuring degree of performance in relation to a common scale across levels, and could assign marks according to both their expectation and the level being studied.

---

[2] Although school marks would almost certainly have assisted teachers in making their assessments of the expected performance of their students on the HSC Examination.

In each course area and for each school there were four sets of scores available for analysis and examination:

(1) School assessment scores, or more accurately teacher expectation scores of performance on the HSC Examination,

(2) Higher School Certificate scores,

(3) ASAT scores,

(4) school assessment scores which were to be rescaled using ASAT as a moderating variable (derived from (1) and (3) above).

While it was possible to consider the data at each of the three levels for each subject, a preliminary examination of the data in this way revealed that the sizes of the groups were, in general, too small for a worthwhile study. Since the combining of scores across levels was a meaningful operation for both the HSC Examination marks and for the school assessment measures, it was decided that the use of a single set of scores for each subject or course area was to be preferred.

The total numbers of students in each of the courses varied considerably, even after combining across levels, and ranged from 1442 students in the subject of English to only 54 students in the subject of Home Science and 26 students in Music. Courses which in total attracted smaller numbers of students were not considered. In general, all 20 schools entered students for the HSC Examination in each subject area, and thus the numbers of schools and students within schools were large enough for effective analysis. However, for Music and Home Science the numbers of schools and of students within schools were small and as a consequence the data presented must be viewed with some misgivings. The relevant numbers of students and schools are recorded in the Tables which follow.

## Computing

Undertaking the work of computing posed few problems with respect to the procedures of rescaling by the Linear Transformation Method. For rescaling by the Equipercentile Method the program described by Lindsay and Prichard (1971) was obtained from the authors, but was found to be unsuitable for the computer system being used. Consequently, it was necessary to develop a program that carried out the equipercentile scaling procedures in a manner corresponding to that described by Flanagan (1951) and that carried out the two steps in rescaling referred to by Lindsay and Prichard (1971:204) of (1) interpolation of the cumulative frequency distributions for a test and its moderator and (2) curve fitting. First, for each raw score for both the school assessment scores (X) and the ASAT scores (Y) the corresponding interpolated score of $Y_n$ and $X_n$ were computed from the cumulative percentage score distributions. In the second operation the two sets of scores,

$$X = \{x_n \mid \text{min. of test } X \leq n \leq \text{max. of test } X\}$$
$$Y = \{y_n \mid \text{min. of test } Y \leq n \leq \text{max. of test } Y\}$$

were successively fitted with polynomials of degree 1 to 4, treating Y as the dependent variable and X as the independent variable. The goodness of fit of each polynomial was tested using a Chi-squared test and the curve which gave the minimum value was taken as the best fit to the data. The third and final step involved substitution of school assessment scores in the fitted polynomial to obtain the rescaled score corresponding to each raw school assessment course score.

Rescaling was carried out by both the linear transformation method and the equipercentile method, and the rescaled scores were rounded to the nearest whole number before calculating correlation coefficients.

## Research Questions

There were several important questions to be asked of the data.

1. Is rescaling of school assessment scores necessary with this set of data?
2. Is ASAT an appropriate reference test for the purposes of rescaling?
3. Is rescaling of the school assessment scores, if necessary and appropriate, best undertaken by the linear transformation method or by the equipercentile method?

These questions led to several propositions for analysis and testing.

1. In justifying the need to rescale it is necessary to show that between school groups, school assessment scores correlate less with scholastic aptitude scores than do HSC scores with scholastic aptitude scores. Thus teachers' judgments of level of performance as related to the level of developed ability of a course group are not as accurate as the HSC assessment assuming ASAT to be a sound criterion of level of developed ability:

$$ER_{YT} \text{ (ASAT X HSC)} > ER_{YX} \text{ (ASAT X SA)}$$

2. In order to assess whether ASAT is an appropriate reference test for the rescaling of school assessment scores it is necessary to examine whether the between student correlation between HSC taken as the criterion and ASAT is high:

$$R_{YT}\ (\text{ASAT X HSC}) >> 0.$$

3. In order that school assessment scores should be seen as a suitable measure of student performance it is important that the between student within school correlation between school assessment scores and HSC scores should be high, assuming that HSC scores form an appropriate criterion:

$$WR_{XT}\ (\text{SA X HSC}) >> 0.$$

4. As an alternative indicator of whether ASAT is an appropriate reference test for the rescaling of school assessment scores it is necessary to examine whether the between student within school correlation between ASAT scores and school assessment scores, taken as the criterion, is high:

$$WR_{YX}\ (\text{ASAT X SA}) >> 0.$$

5. It is important to examine whether the between school correlation between ASAT scores and HSC scores is high in order to ensure that the ranking of schools would be similar under rescaling as under HSC:

$$ER_{YT}\ (\text{ASAT X HSC}) >> 0.$$

6. In addition it is necessary to show that the between school correlation between ASAT scores and HSC scores is greater than the between school correlation between school assessment scores and HSC scores, to guarantee that schools would not suffer in the rescaling process relative to their performance on the HSC:

$$ER_{YT}\ (\text{ASAT X HSC}) > ER_{XT}\ (\text{SA X HSC})$$

7. It is necessary to show that the rescaled scores are strongly related to HSC performance. This requires that the between student correlation between HSC as a criterion and the rescaled scores is high:

$$R_{ZT}\ (\text{RSA X HSC}) >> 0.$$

8. In order to select between the two methods of rescaling it is necessary to calculate the between student correlation between the rescaled scores and the HSC scores for the two methods of rescaling scores. It was initially supposed that the equipercentile scaling method would yield rescaled scores which would correlate more highly with HSC scores than would the rescaled scores obtained by the linear transformation method:

$$R_{ZT}\ \underset{\text{LINEAR}}{(\text{RSA X HSC})} < R_{ZT}\underset{\text{EQUIPERCENTILE}}{(\text{RSA X HSC})}$$

**Results**

In the presentation of results all three correlations, the between student correlation (R), the between school correlation (ER), and the between student within school correlation (WR) between the two measures have been recorded in order to give full evidence of the relationships between the different measures. The correlations have been recorded to two decimal places for all 13 subjects for which analyses were carried out, and the median correlation for each set has been given. In addition, the numbers of students and the number of schools included in the analyses together with the range of students within schools has been presented to provide some evidence for the assessment of the stability of the correlation coefficients calculated.

**Proposition 1**

$$ER_{YT}\ (\text{ASAT X HSC}) > ER_{YX}\ (\text{ASAT X SA})$$

The evidence for the examination of this proposition is recorded in Columns 2 and 5 of Table 1. In examining the relationships between the between school correlation for ASAT scores with both the Higher School Certificate and the school assessment scores it should be noted that in all cases except two, Ancient History and Art, the between school correlations are less for school assessment scores with ASAT than for the Higher School Certificate Scores with ASAT, sometimes substantially so. It is therefore argued that the teachers' global judgment of the level of scholastic aptitude of their course group is less accurate than indicated by the HSC. Thus gains would be made by using the HSC scores in preference to unscaled school assessment scores since the former reflect more closely level of developed ability. This of course assumes that ASAT, as a measure of developed ability, is a sound criterion.

**Proposition 2**

$$R_{YT}\ (\text{ASAT X HSC}) >> 0.$$

The data for the examination of Proposition 2 that the between student correlation between scores on the Australian

Table 1. Correlations between ASAT Scores and Higher School Certificate Scores and School Assessment Scores for ACT 1975

| Correlations[a] Subject | | (HSC X ASAT) $R_{YT}$ | $ER_{YT}$ | $WR_{YT}$ | (SA X ASAT) $R_{YX}$ | $ER_{YX}$ | $WR_{YX}$ | Numbers Total Students | Schools | Range of Students in Schools |
|---|---|---|---|---|---|---|---|---|---|---|
| English | | 51 | 70 | 50 | 54 | 43 | 55 | 1442 | 20 | 30 - 111 |
| Mathematics | | 64 | 57 | 64 | 61 | 49 | 62 | 1290 | 20 | 30 - 101 |
| Science | | 69 | 73 | 69 | 64 | 68 | 63 | 1241 | 20 | 23 - 94 |
| Modern History | | 57 | 71 | 56 | 63 | 58 | 63 | 636 | 20 | 13 - 47 |
| Ancient History | | 47 | 53 | 45 | 52 | 65 | 49 | 317 | 19 | 4 - 34 |
| Geography | | 50 | 56 | 49 | 52 | 43 | 54 | 862 | 20 | 18 - 76 |
| Economics | | 64 | 58 | 64 | 60 | 29 | 65 | 679 | 20 | 19 - 51 |
| French | | 55 | 14 | 65 | 50 | 6 | 60 | 168 | 17 | 3 - 24 |
| German | | 35 | 43 | 31 | 23 | -12 | 39 | 83 | 15 | 3 - 13 |
| Music | | 45 | 98 | 43 | 38 | 47 | 38 | 26 | 3 | 3 - 14 |
| Art | | 39 | -9 | 49 | 41 | 30 | 44 | 156 | 16 | 5 - 21 |
| Industrial Art | | 67 | 55 | 71 | 46 | 6 | 57 | 145 | 14 | 3 - 16 |
| Home Science | | 51 | 63 | 47 | 53 | 37 | 65 | 54 | 8 | 4 - 11 |
| Median | | 55 | 57 | 50 | 52 | 43 | 57 | 317 | 19 | 30 |
| Range | Low | 35 | -9 | 31 | 23 | -12 | 38 | 26 | 3 | 7 |
| | High | 69 | 98 | 71 | 64 | 68 | 65 | 1442 | 20 | 81 |

[a] Correlations recorded with decimal points omitted.

Scholastic Aptitude Test and scores on the Higher School Certificate are high are also recorded in Column 1 of Table 1. The predictive power of ASAT for performance on the HSC ranges from a low in German of 35 to a high in Science of 69 with a median value of 55. While these correlations are at best moderate, they indicate that performance on the Higher School Certificate subjects are associated with a factor of general scholastic aptitude and have sufficient in common with ASAT, and presumably with each other, to warrant the combining of subject scores to form a total score as has been the practice for many years past. It is of course possible that performance on the different subjects is even more strongly dependent on a factor of general intellectual ability than is indicated by the correlations recorded between HSC scores and ASAT, which is believed to assess developed ability or scholastic aptitude. We have, however, no further evidence of this.

Proposition 3

$$WR_{XT} \text{ (SA X HSC) } >> 0.$$

The data for the examination of Proposition 3 are given in Column 3 of Table 2 where the between student within school correlation between school assessment scores and Higher School Certificate scores are recorded. Without exception teachers within schools are well able to assess the relative performance of their students on HSC since the correlations range from a low of 0.77 to a high of 0.93 with a median correlation of 0.84. It should also be noted from Column 1 in Table 2 that teachers are also reasonably good predictors of student performance on HSC since the between student correlations range from a low of 0.68 to a high of 0.92 with a median value of 0.80. Nevertheless, it would be unsound to assume that when the HSC is removed the teachers will be able to assess performance relative to the total group of students with the same degree of accuracy, although it is to be anticipated that they will continue to be able to assess accurately student performance within the school. The accuracy of their present level of expectation, relative to performance on the Higher School Certificate, is no doubt linked with tightly defined syllabuses associated with the HSC, and the constant and standard style of examination paper used by many HSC examiners.

Proposition 4

$$WR_{YX} \text{ (ASAT X SA) } >> 0.$$

A further indicator of the meaningfulness of ASAT as a reference variable for the rescaling of school assessment scores would be that the between student within school correlations between ASAT and the school assessment scores were high. The data that will permit an examination of this proposition are recorded in Column 6 of Table 1. The evidence indicates that, within schools, ASAT is sufficiently highly correlated with school

Table 2. Correlations between Higher School Certificate Scores and School Assessment and Rescaled[a] School Assessment Scores for ACT 1975

| Correlation[b] | (SA X HSC) | | | (RSA X HSC) | | | Numbers | | |
|---|---|---|---|---|---|---|---|---|---|
| Subject | $R_{XT}$ | $ER_{XT}$ | $WR_{XT}$ | $R_{ZT}$ | $ER_{ZT}$ | $WR_{ZT}$ | Total Students | Schools | Range of Students in Schools |
| English | 73 | 43 | 76 | 75 | 70 | 75 | 1442 | 20 | 30 - 111 |
| Mathematics | 92 | 80 | 93 | 90 | 55 | 93 | 1290 | 20 | 30 - 101 |
| Science | 91 | 87 | 92 | 90 | 75 | 92 | 1241 | 20 | 23 - 94 |
| Modern History | 81 | 78 | 82 | 80 | 68 | 82 | 636 | 20 | 13 - 47 |
| Ancient History | 73 | 51 | 78 | 72 | 55 | 77 | 317 | 19 | 4 - 34 |
| Geography | 76 | 61 | 78 | 75 | 56 | 78 | 862 | 20 | 18 - 76 |
| Economics | 84 | 83 | 84 | 82 | 59 | 84 | 679 | 20 | 19 - 51 |
| French | 84 | 74 | 87 | 73 | 16 | 88 | 168 | 17 | 3 - 24 |
| German | 83 | 73 | 89 | 65 | 43 | 76 | 83 | 15 | 3 - 13 |
| Music | 80 | 65 | 91 | 87 | 98 | 89 | 26 | 3 | 3 - 14 |
| Art | 73 | 51 | 78 | 62 | -14 | 76 | 156 | 16 | 5 - 21 |
| Industrial Art | 75 | 59 | 79 | 71 | 53 | 76 | 145 | 14 | 3 - 16 |
| Home Science | 68 | 61 | 77 | 68 | 66 | 70 | 54 | 8 | 4 - 11 |
| Median | 80 | 73 | 84 | 75 | 56 | 78 | 317 | 19 | 30 |
| Range Low | 68 | 51 | 77 | 62 | -14 | 70 | 26 | 3 | 7 |
| High | 92 | 87 | 93 | 90 | 98 | 93 | 1442 | 20 | 81 |

[a] Rescaling undertaken by Linear Transformation Method.

[b] Correlations recorded with decimal points omitted.

assessment measures for each subject, to support its use as a moderating variable. The correlations range from a low of 0.38 for Music to a high of 0.65 for Economics and Home Science with a median value of 0.57 for Industrial Art. Clearly performance assessed at school is moderately related to an underlying factor of scholastic aptitude.

Proposition 5

$$ER_{YT} \text{ (ASAT X HSC)} >> 0.$$

The evidence for the examination of Proposition 5 are recorded in Column 2 of Table 1. This proposition states that the between school correlation between school mean scores on the Australian Scholastic Aptitude Test and school mean scores on the HSC in each subject should be high. In general, the correlations are high, but for French a value of 0.14 and for Art a negative correlation of -0.09 are recorded. There are, however, surprisingly high correlations for English of 0.70 and for Music of 0.98. The median correlation of 0.57 is also recorded for Mathematics. Clearly in most subjects there will not be a marked change in the ranking of schools relative to their mean performance on the HSC after rescaling with ASAT. These results are of some concern, for while they indicate that for most subjects ASAT is a suitable reference test, it is clearly not appropriate for French and for Art. It should also be noted that while the between student correlations between HSC scores and ASAT scores for these two subjects are not high they are not the lowest recorded (see Column 1 of Table 1).

Proposition 6

$$ER_{YT} \text{ (ASAT X HSC)} > ER_{XT} \text{ (SA X HSC)}$$

In the light of the findings of Proposition 5 it becomes important to examine whether the between school correlations between HSC scores and ASAT scores are greater than the between school correlation between school assessment scores and HSC scores. If they are less then it might be argued that school groups have suffered in the operation of rescaling. These relevant data are recorded in Column 2 of Table 1 and Column 2 of Table 2 and in order that this proposition may be subjected to closer scrutiny and its implications more fully assessed, Table 3 presents these two sets of data and records, in addition, the difference between these two correlations ($ER_{yt}$ - $ER_{xt}$).

The evidence recorded in Table 3 clearly indicates that in the rescaling of the school assessment scores, the school means for school assessment scores which before rescaling were quite strongly correlated with the school means on the HSC, will for most courses, after rescaling, be less strongly correlated with the HSC school means. In general, the relationship has been reduced as indicated by the negative differences recorded in Table 3, but in English and Music there is substantial reordering of school means which are now more strongly related to the HSC

Table 3. Between School Correlations between both Higher School Certificate Scores and School Assessment Scores and ASAT Scores for ACT, 1975

| Subject | | Between School Correlations[a] ER (ASAT X HSC) | (SA X HSC) | Difference $ER_{YT}-ER_{XT}$ | Number of Schools |
|---|---|---|---|---|---|
| English | | 70 | 43 | 27 | 20 |
| Mathematics | | 57 | 80 | -23 | 20 |
| Science | | 73 | 87 | -14 | 20 |
| Modern History | | 71 | 78 | -7 | 20 |
| Ancient History | | 53 | 51 | 2 | 19 |
| Geography | | 56 | 61 | -5 | 20 |
| Economics | | 58 | 83 | -25 | 20 |
| French | | 14 | 74 | -60 | 17 |
| German | | 43 | 73 | -30 | 15 |
| Music | | 98 | 65 | 33 | 3 |
| Art | | -9 | 51 | -60 | 16 |
| Industrial Art | | 55 | 59 | -4 | 14 |
| Home Science | | 63 | 61 | 2 | 8 |
| Median | | 57 | 65 | -7 | 19 |
| Range | Low | -9 | 43 | -60 | 3 |
| | High | 98 | 87 | 30 | 20 |

[a] Correlations recorded with decimal points omitted.

scores than previously. Yet it must be recognized that with the exception of French and Art the between school correlations between the ASAT scores and the HSC scores are not low, and the school mean of the rescaled school assessment will hold approximately the same correlation with the mean of the HSC scores as did the unscaled school assessments. It must be clearly recognized that in bringing all subjects to a common scale by using ASAT as a reference test some reordering of schools takes place, and that this reordering does not follow the ranking of the school means on HSC but diverges from this ranking. This raises the question whether HSC forms an appropriate criterion for judging the effectiveness of rescaling, and whether injustice has been done in the past in combining HSC scores to form a total score.

In addition, it should be noted that the school assessments were made by teachers who were preparing students for the HSC Examination with the same prescriptive syllabus for each school and college. As mentioned earlier, when the external examination is gone, teachers will not have the same yard-stick on which to place the school mean. They should, however, still be able to assess the relative merit of students within their schools. Thus, the need to rescale to an objective yard-stick (ASAT) will increase.

Proposition 7

$$R_{ZT}\ (RSA\ X\ HSC) >> 0.$$

In proposition 7 the magnitude of the between student correlation between the rescaled school assessment scores and the Higher School Certificate scores is examined for each course. The data for the linear transformation method of rescaling are recorded in Column 1 of Table 4. The correlations are satisfactorily high. They range from a low for Art, as might be expected, of 0.62, to a high for Mathematics and Science of 0.90, and with a median value of 0.75. Nevertheless, these correlations are slightly lower than the within school correlations between the same measures which are recorded in Column 3 of Table 4. In no subject are they higher, but this is not surprising, since it might be expected that, within schools, the teachers would be able to estimate the performance of their students on HSC rather better than would be accomplished by rescaling process.

Although serious problems might have been anticipated both for French and for Art these are not critical, presumably because the between school variance was small with respect to the within school variance on both the school assessment scores and on the ASAT scores, and the correlation ratios for these variables were also low. As a consequence the schools were reordered by the rescaling operation, but not to such an extent that rescaled scores were no longer quite strongly correlated with the Higher School Certificate scores.

**Table 4.** Correlations between Rescaled School Assessment Scores and Higher School Certificate Scores for both Linear Transformation Method and Equipercentile Method for ACT, 1975

| Correlations[a] | Linear Transformation | | | Equipercentile | | | Difference | Number of Students |
|---|---|---|---|---|---|---|---|---|
| Subject | $R^L_{ZT}$ | $ER_{ZT}$ | $WR_{ZT}$ | $R^E_{ZT}$ | $ER_{ZT}$ | $WR_{ZT}$ | $R^L_{ZT} - R^E_{ZT}$ | |
| English | 75 | 70 | 75 | 71 | 67 | 72 | 4 | 1442 |
| Mathematics | 90 | 55 | 93 | 87 | 56 | 90 | 3 | 1290 |
| Science | 90 | 75 | 92 | 88 | 74 | 90 | 2 | 1241 |
| Modern History | 80 | 68 | 82 | 80 | 68 | 82 | 0 | 636 |
| Ancient History | 72 | 55 | 77 | 72 | 51 | 77 | 0 | 317 |
| Geography | 75 | 56 | 78 | 76 | 55 | 78 | -1 | 862 |
| Economics | 82 | 59 | 84 | 81 | 56 | 84 | 1 | 679 |
| French | 73 | 16 | 88 | 72 | 12 | 87 | 1 | 168 |
| German | 65 | 43 | 76 | 62 | 45 | 71 | 3 | 83 |
| Music | 87 | 98 | 89 | 84 | 93 | 85 | 3 | 26 |
| Art | 62 | -14 | 76 | 60 | -14 | 75 | 2 | 156 |
| Industrial Art | 71 | 53 | 76 | 70 | 59 | 74 | 1 | 145 |
| Home Science | 68 | 66 | 70 | 62 | 64 | 62 | 6 | 54 |
| Median | 75 | 56 | 78 | 72 | 56 | 78 | 2 | - |
| Range Low | 62 | -14 | 70 | 62 | -14 | 62 | -1 | - |
| High | 90 | 98 | 93 | 88 | 93 | 90 | 6 | - |

[a] Correlations recorded with decimal points omitted.

**Proposition 8**

$$R_{ZT}\,\underset{\text{LINEAR}}{(\text{RSA X HSC})} < R_{ZT}\,\underset{\text{EQUIPERCENTILE}}{(\text{RSA X HSC})}$$

In this proposition we compare the methods of linear rescaling and equipercentile rescaling. Because the equipercentile method would take account both of the shape of the school assessment score distribution and that of the ASAT score distribution it was believed to be more appropriate than the linear transformation method. Consequently it was assumed initially that the rescaled course scores obtained by the equipercentile scaling method would be more strongly correlated with the HSC scores than would the rescaled scores obtained by the linear transformation method. The values of the between student correlations under examination are recorded in Columns 1 and 4 of Table 4. The difference between the correlations associated with the linear rescaling method and those associated with the equipercentile rescaling method are recorded in Column 7. In all cases except one, the linear rescaling method produced scores with marginally higher correlations with HSC scores than did the equipercentile rescaling method. The differences ranged from a low of -0.01 to a high of 0.06 with a median difference of 0.02.

It is necessary to ask why the rescaled scores obtained by the linear method should correlate more highly with the HSC scores than do those obtained by the equipercentile method. It should be recalled that the linear rescaling method does not change the shape of the teacher assessment score distribution during rescaling, whereas the equipercentile method does. It would appear that the teacher assessment score distributions, in general, follow more closely the HSC score distributions than do the ASAT score distributions which are incorporated into rescaled scores by the equipercentile method. Thus there is a stronger relationship between the scores rescaled by linear rescaling and the HSC scores than there is between the scores obtained by equipercentile scaling and the HSC scores.

If it is assumed that the teacher assessment scores reflect more accurately a student's worth, then the linear rescaling is to be preferred. Yet, if it is considered desirable that the rescaled scores should reflect both teacher assessment and the underlying developed ability or scholastic aptitude of the course group, then the equipercentile method is to be preferred. Moreover, it could again be asked whether the HSC course scores form an appropriate criterion. Perhaps the choice between the two methods should not be resolved on either empirical or logical grounds, but rather on the basis of parsimony and simplicity. Since the linear transformation method is simpler to operate and clearly does not disadvantage the student group relative to the former Higher School Certificate, it may be argued that it is the more appropriate procedure to employ.

## Conclusion

In this paper we have asked three important questions concerned with whether rescaling is necessary, whether ASAT is an appropriate reference test, and whether the linear transformation method or the equipercentile method of rescaling is to be preferred. In providing answers to these questions it is essential to recognize that a policy decision was made nearly three years ago for schools in the ACT to break from the New South Wales Higher School Certificate program, and to withdraw students in the Territory from entry for the NSW Higher School Certificate Examination. Thus the colleges moved from a situation with common courses, as has largely continued to exist in Queensland, to one where a great variety of courses were being taught to students of differing levels of ability. Under these circumstances the rescaling of teachers marks and assessments was unavoidable if Tertiary Entry Scores were to be calculated. The data collected in the 1975 trial study, which involved teacher expectations of student performance on a common examination, were in some respects inappropriate for an adequate testing of all the relevant research questions that might have been asked. Nevertheless, they were the only data available and served to provide an opportunity for a study of some of the consequences of the proposed rescaling operations.

The second question was concerned with whether ASAT was an appropriate reference test. This paper has only considered whether ASAT could be used effectively for rescaling teacher assessment scores from different colleges and high schools within the different subject areas. The evidence presented indicated that the procedures were, in general, highly satisfactory, but shortcomings were observed for modern languages, particularly French, and for Art. In part, the effectiveness of the rescaling procedures was linked with low correlation ratios, for example, 0.069 for English and 0.086 for ASAT. As a consequence, since the school means were relatively homogeneous, marked changes did not occur in the rescaling operation when adjustments were made for differences between schools within the one subject area. More substantial changes were likely to occur when the scores in different subject areas were brought to a common scale. A further study of such effects has been undertaken using the same data by Mr D.E. Morgan at the Canberra College of Advanced Education.

The reasons why the rescaling procedures were less effective for French and Art arose most probably because the students taking these courses performed poorly on the quantitative aspects of the ASAT reference test. This would seem to indicate that the developed abilities of the students who were taking these subjects differed from those required by the ASAT test. This relationship should not be taken as a criticism of either the ASAT test or the students taking such subjects. The evidence merely establishes that the attempt to bring scores in the different subjects to a common scale was less effective in some subject areas than in others.

It is important to note that this is a criticism of the use of a total score in selection for tertiary entry and not of the ASAT test or of the rescaling procedure employed. The members of the Agency have been well aware of such problems from the outset and have advocated the use of the student profile for selection purposes rather than the use of a total score.

The third question examined in this paper has been concerned with whether the linear transformation method or the equipercentile method of rescaling was the more appropriate. The differences between the two procedures were largely slight. A third procedure was developed in the course of the investigation by which the shape of the ASAT score distribution determined the final distribution of rescaled school assessment scores. The results obtained with this procedure were only marginally different from those resulting from the other two methods. However, it was shown that scores rescaled in this way diverged further from the HSC score distribution than where either of the other two procedures was employed. The question remains as to whether the teacher assessment scores provide the most valid shape of the underlying score distribution, whether the scholastic aptitude as measured by ASAT would be more valid underlying distribution, or whether a compromise between the two as calculated by the equipercentile procedure would be the most appropriate. The question is a theoretical one since the empirical evidence has indicated that the differences were not large. Perhaps, however, the simplest procedure which involves the linear transformation method of rescaling should be preferred, and this procedure is being used in 1977. Nevertheless, it will be necessary to examine carefully what happens to the scores of individuals in the rescaling operation particularly in the tails of the distributions in order to determine whether certain gains would be made if an alternative method was used.

# REFERENCES

Bell, R.C. 1977 *A Psychometric Study of the Australian Scholastic Aptitude Test (Series B)*. Nedlands, Western Australia: University of Western Australia.

Brennan, M.H., (Chairman). 1975 Report of ASAT Committee. Unpublished Report issued by The Public Examinations Board of South Australia. Adelaide, South Australia.

Campbell, R.J. (Chairman). 1973 *Secondary Education for Canberra: Report of the Working Committee on College Proposals for the Australian Capital Territory*. Canberra: Australian Government Publishing Service.

Cook, J.S. and Cooney, G.H. 1976 The Scaling of Marks at the Higher School Certificate Examination. Unpublished paper issued by the New South Wales Department of Education, Sydney.

Cooney, G.H. 1975 'Standardization Procedures Involving Moderator Variables - Some Theoretical Considerations'. *Australian Journal of Education*, 19(1):50-63.

Currie, G. (Chairman). 1967 *An Independent Education Authority for the Australian Capital Territory*. *Report of a Working Party*. Canberra: Department of Adult Education, Australian National University.

Flanagan, J.C. 1951 'Units, Scores and Norms'. In E.F. Lindquist (ed.) *Educational Measurement*. Washington, D.C.: American Council on Education, pp 752-760.

Hannan, M.T. 1971 *Aggregation and Disaggregation in Sociology*. Lexington, Mass.: Lexington Books, D.C. Heath and Company.

Howard, M. 1958 'The conversion of scores to a uniform scale'. *British Journal of Statistical Psychology*: 11 (Part II): 30-40.

Lindsay, C. and Prichard, M.A. 1971 'An Analytical Procedure for the Equipercentile Method of Equating Tests'. *Journal of Educational Measurement*, 8(3):203-208.

McGaw, B. 1975 The use of Rescaled Teacher Assessments in Selection for Tertiary Study. Paper read at Symposium on 'The Prediction of Student Success in Tertiary Institutions'. Australian Psychological Association Conference. Melbourne.

McGaw, B. Warry, R. and McBryde, B. 1975 *The Queensland Grade 12 Study*: *Report No 2. Validation of Aptitude Measures for the Rescaling of School Assessments*. Brisbane: Research Branch, Department of Education, Queensland.

Morgan, G. 1977 Report on the 1976 Australian Scholastic Aptitude Test. Unpublished paper issued by the ACER to test users.

Robinson, W.S. 1950 'Ecological Correlations and the Behaviour of Individuals'. *American Sociological Review*, 15: 351-357.

Thomson, P.P. 1974 *Assessment for Australian Capital Territory Secondary Schools*. Hawthorn, Victoria: ACER.

## SATURDAY 12 NOVEMBER

## 11.00 - 11.45am

| PAPERS | Rooms |
|---|---|
| D. Cohen, M. Harrison (Macquarie University)<br>Perceptions of Decision-making in School Based Curriculum Development (Curriculum Action Project) | Ladies Drawing Room |
| P. Fensham, L. Ingvarson (Monash University)<br>OMITTED Ethnographic Evaluation of Science Classrooms | Seminar Room A |
| J.M. Atkinson, G.R. Trevilco (Rusden State College)<br>Evaluation of School Experience in Pre-Service Teacher Education | Seminar Room B |
| WORKSHOP ON ETHNIC STUDIES | Meeting Room |
| G. Rowley (La Trobe University)<br>How Change is Measured and Why It Usually Isn't: Implications for Educational Research | Gateway Room |
| [illegible] Rowlands (V.U.S.E.B.)<br>Some Problems of Curriculum Evaluation at Grade 12 Level | Seminar Room D |

# CURRICULUM ACTION PROJECT

A JOINT STUDY FUNDED BY THE EDUCATION RESEARCH AND DEVELOPMENT COMMITTEE CONDUCTED BY MACQUARIE UNIVERSITY AND THE UNIVERSITY OF QUEENSLAND IN COOPERATION WITH THE CURRICULUM DEVELOPMENT CENTRE

## Perceptions of Decision Making in School Based Curriculum Development

### (Preliminary Findings based on Twelve Schools)

by

David Cohen & Marelle Harrison
Macquarie University

Project Report No 1: October 1977

Research Reports on Curriculum Action Project

No 1 Perceptions of Decision Making
in School Based Curriculum Development
(Preliminary Findings based on Twelve Schools)

(Note: This is the first of a series of Research Reports to be released by the Macquarie University Curriculum Action Project Research Team)

## CONTENTS

| | | Page |
|---|---|---|
| Section I | Introduction | 1 |
| Section II | Curriculum Action Project: Objectives and Methodology | 1 |
| Section III | Findings | 4 |
| SECTION IV | Discussion of Findings | 7 |

## Section I

## INTRODUCTION

The Curriculum Action Project is a national study of curriculum decision-making in Australian junior secondary schools. It was initiated in December 1976 by a research team at Macquarie University which shared with the University of Queensland a $30,000 research grant provided by the Education Research and Development Committee. The proposal also received the support of the Curriculum Development Centre (CDC).

The importance of this study in the contemporary Australian education scene is derived from several contexts. Firstly, there is general concern about the perceived crisis at junior secondary school level with respect to curricula. This concern was expressed in the Triennial Program: 1977-79 of the CDC, entitled "New Directions in Curriculum Development, With an Emphasis on Lower Secondary Schooling". In its report, CDC designated that "the substantive area in which development work should be concentrated during the 1977-79 triennium (is) lower secondary education, i.e., the first four years of secondary schooling" (CDC, 1977: 8).

Secondly, and crucially, there is the purported advent of curriculum "autonomy" or "school-based curriculum development", as announced in various statements in all Australian states in the last decade. For example, in South Australia, the then Director-General of Education A.W. Jones, issued a"Memorandum to Heads of Departmental Schools: Freedom and Authority in the Schools" which in part stated:

> " . . . you have the widest liberty to vary courses, to alter the timetable, to decide the organisation of the school and government within the school, to experiment with teaching methods, assessment of student achievement and in extra-curricular activities."

There has been a long-held perception in Australia that our schools are unlike those in USA and UK, where it has been assumed that autonomy has been granted and exercised. Yet, a recent seminal report of the US NIE Curriculum Development Task Force underscores the

> "feeling of impotence, of not being heard and of having only a limited scope of influence. There is also the view, rather widely held, that somebody else controls the ball game be it professionals, bureaucrats, or some ubiquitous 'they'." (Schaffarzick, 1976: 4)

This generalization was based upon widespread interviews and document analysis. In other words, the perceptions that curricula in schools are controlled locally or that curriculum autonomy is exercised in USA, should now come under critical scrutiny.

What have been the effects of Australian edicts granting autonomy? One effect has been the generation of widespread curriculum activities, as currently being documented in "Curriculum Developments in Australia: Descriptive Profiles" (a national survey commissioned by CDC being completed in 1977 at Macquarie University). Yet, evidence is accruing also of the need for more support within schools if full advantage is to be gained from these newly-granted opportunities to tailor curricula to their specific groups of students. One particular aspect of the support needed is to help them reflect upon their own curriculum decision-making procedures.

## Section II

## CURRICULUM ACTION PROJECT: OBJECTIVES AND METHODOLOGY

The major objective of the Curriculum Action Project is to gather data as the basis for exploring the decision-making processes as they take place in schools, with particular interest in classes at Years 7-10 levels. The total study comprises two major thrusts, namely, a survey and a case study. This present report is confined to the survey. For this survey, a stratified random sample (probability proportional to size) comprising twenty schools in each state and ten in the territories, was drawn for Macquarie University by the Survey Section of the Australian Council for Educational Research. For Education Department schools, permission was sought and readily received from each State Director-General of Education. The schools were then invited to participate, and principals were requested to supply lists of names of staff, indicating the names of those teaching classes at Years 7-10 levels and designating also the names of Heads of Department.

From these lists, the names of ten teachers and three Heads of Department were randomly selected from each participating school. The principal, heads of department and teachers were then each sent a numbered questionnaire together with a personalised letter inviting participation. Participation required completion of a detailed questionnaire, which trials indicated would take between one and two hours to complete.

Several methods were adopted in an attempt to bolster response rates and encourage invitees to accept invitations to participate. In the first place, a media campaign was mounted throughout Australia to highlight the potential importance of the findings. A specially-prepared media release was mailed to press, radio and television outlets in capital cities and in catchment areas of selected schools. Secondly, because of the time (1-2 hours) required for participation by respondents, each school principal was requested to attempt to find released time for invitees or to otherwise compensate participants where this was feasible. Teachers in a parallel United States of America study were paid honoraria of $25 upon the return of completed questionnaires.

Although separate questionnaires were necessarily developed for the three groups of invitees, they contained many items common to all three forms. The appropriate numbered questionnaire (Form A (white) for teachers, Form B (hellow) for heads of department, and Form C (pink) for principals) was included with each invitation, together with a reply-paid addressed envelope for return of the completed questionnaire. Invitees were requested to respond by a designated date three weeks from the mailing date to each invitee. They were assured both of the confidentiality of the data, and of the provision of feedback to all participating schools.

As Table 1 illustrates, there has been a very good response to participation in this study from schools and invitees. The response rate is, in a sense, one index of commitment to improving curriculum in schools.

Table 1: Numbers of Responses to CAP Questionnaires, by States and by Responsibilities

| STATE | Number of Schools Initially Initiated | Number of School Acceptances for Participation (as at 30/9/77) | Number of Invitations Mailed (30/9/77) | | | Number of Respondents (30/9/77) | | |
|---|---|---|---|---|---|---|---|---|
| | | | Principals | Heads of Dept. | Teachers | Principals | Heads of Dept. | Teachers |
| N.S.W. | 20 | 16<br>(80%) | 16 | 48 | 160 | 9<br>(56%) | 31<br>(65%) | 62<br>(39%) |
| Victoria | 20 | 18<br>(90%) | 17 | 51 | 170 | 11<br>(65%) | 14<br>(27%) | 57<br>(33%) |
| Queensland | 20 | 14<br>(70%) | 14 | 42 | 140 | 8<br>(57%) | 15<br>(36%) | 35<br>(25%) |
| South Australia | 20 | 16<br>(80%) | 16 | 48 | 160 | 12<br>(75%) | 18<br>(36%) | 61<br>(38%) |
| Western Australia | 18 | 12<br>(67%) | 12 | 36 | 120 | 8<br>(67%) | 16<br>(44%) | 45<br>(38%) |
| Tasmania | 19 | 17<br>(89%) | 17 | 51 | 170 | 11<br>(65%) | 21<br>(41%) | 46<br>(21%) |
| A.C.T. | 9 | 40<br>(44%) | 4 | 12 | 40 | - | - | 2 |
| N.T. | 6 | 2<br>(33%) | 2 | 6 | 20 | 1<br>(50%) | 1<br>(17%) | 2<br>(10%) |
| TOTALS | 132 | 99<br>(75%) | 98 | 294<br>1372 | 980 | 60<br>(61%) | 115<br>(39%)<br>485<br>35.35% | 310<br>(31%) |

The questionnaires were devised, trialled and modified on the basis of the trials. They were designed to elicit data concerning many aspects of curriculum decision-making within schools, departments and classrooms. Emphases were placed upon such concerns as:

What curriculum decisions are perceived as being made?

How are these curriculum decisions perceived as being made? (i.e. what strategies, arrangements and procedures exist for meetings and decision-making? What criteria (if any) are employed?)

What factors - human and physical, internal and external - are perceived as influencing (facilitating or inhibiting) decision-making in curriculum?

Who shares in the curriculum decision-making procedures? (internal and external participants).

Items were designed to probe data concerning decision-making processes and participants, in relation to curriculum concerns such as objectives, content, teaching methods, materials, sequencing, scope and interrelationships, and student evaluation. The Curriculum Action Model (Figure 1) indicates the range of dimensions and interrelationships explored by questions contained in the questionnaires. Of the 720 theoretical possibilities generated by the model, questions were devised to tap at least 440 cells, providing a degree of construct validity to the study.

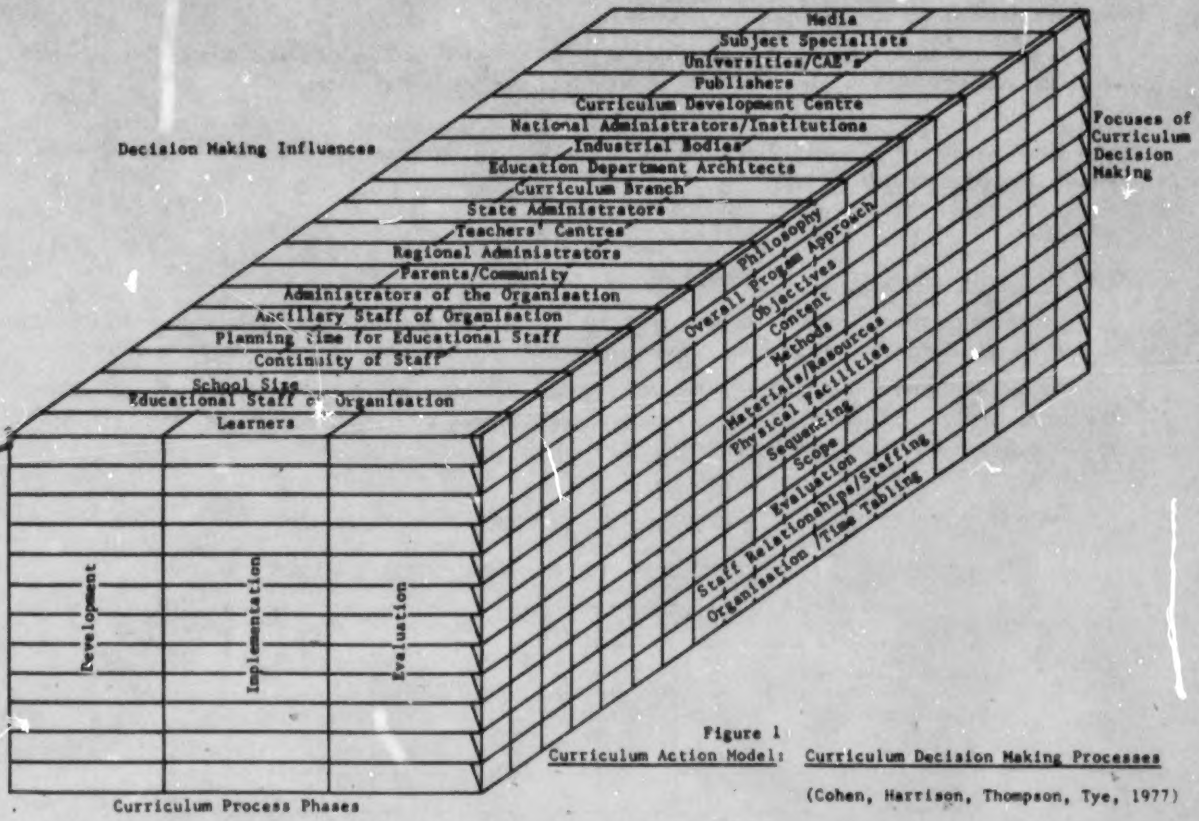

Media
Subject Specialists
Universities/CAE's
Publishers
Curriculum Development Centre
National Administrators/Institutions
Industrial Bodies
Education Department Architects
Curriculum Branch
State Administrators
Teachers' Centres
Regional Administrators
Parents/Community
Administrators of the Organisation
Ancillary Staff of Organisation
Planning time for Educational Staff
Continuity of Staff
School Size
Educational Staff of Organisation
Learners
Decision Making Influences
Focuses of Curriculum Decision Making
Philosophy
Overall Program Approach
Objectives
Content
Methods
Materials/Resources
Physical Facilities
Sequencing
Scope
Evaluation
Staff Relationships/Staffing
Organisation/Time Tabling
Development
Implementation
Evaluation
Curriculum Process Phases

Figure 1

Curriculum Action Model: Curriculum Decision Making Processes

(Cohen, Harrison, Thompson, Tye, 1977)

Our initial research concerns were to make hunches about such "variables" as:

(a) perceptions of participants in curriculum decision-making;

(b) effects of school climate indices, such as perceived communication opportunities, and relationships between staff members;

(c) perceptions of curriculum autonomy/control as related to possible state influences;

(d) school size;

(e) innovativeness with respect to curriculum;

(f) influence of resources (time, space and people).

The parameters of and procedures for the study were deliberately flexible in order to allow their modification on the bases of feedback progressively received from responses to questionnaires and via participant-observations.

For the purposes of this first interim report, data reported are restricted to those from a pilot study comprising respondents from two schools in each state. Selection of these schools was based upon:

(1) the two schools being "different" from each other (with respect to size, location, and/or government/non-government);

(2) a minimum response, at the time of selection, of the principal, two departmental heads and five other teachers.

It is important to note the following characteristics of the data:

(i) they represent self-reports by respondents of their perceptions of events and people, and so are inevitably influenced by the insights and values of respondents;

(ii) they are related specifically to curriculum decision-making (formal and informal) whereas data was not specifically sought concerning other aspects of classroom activities;

(iii) the data reported here are pilot data from twelve Australian secondary schools: it is based upon an analysis of 99 cases, or about 7% of the total sample. It cannot therefore be assumed that the findings are:

(a) generalisable to any other secondary schools;

(b) generalisable to any Year levels other than Years 7-10 (for example, to primary schools).

Thus, this report is based on preliminary and exploratory conceptualization, and may illuminate some trends which assist in the formulation of hypotheses. However, at the best, this preliminary report provides impressionistic rather than generalisable findings.

It should be emphasised that it is not intended that the research team should make judgemental statements about the data provided - the concern is with actual rather than "desirable" procedures.

The supplementary data being gathered in the case-study from observations, interviews and participations in a high school setting in relation to curriculum decision-making, are adding additional in-depth dimensions to this presently-reported study.

## SECTION III: FINDINGS

### 1.0 Autonomy in Curriculum Decision-making

The central concern of this preliminary report has been with the question of curriculum autonomy reflected through the question of who is perceived as exercising authority in the making of curriculum decisions.

The preliminary data from the Curriculum Action Project at Macquarie University appear to suggest that, in the context of Years 7-10 of the Australian secondary school, there are marked interstate differences with respect to perceptions of curriculum autonomy (see Table 2).

Table 2: Ranked Perceptions of Authority in Curriculum Decision-making by State

| Authority | STATE | | | | | | TOTAL |
|---|---|---|---|---|---|---|---|
| | NSW | VIC | QUE | SA | WA | TAS | |
| Head of Department | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| Deputy principal | 6 | 3 | 4 | 3 | 8 | 8 | 5 |
| Principal | 7 | 4 | 2 | 1 | 4 | 3 | 6 |
| Individual parents | 11 | 10 | 9 | 9 | 10 | 11 | 11 |
| P & C | 10 | 10 | 10 | 10 | 11 | 9 | 10 |
| CDC | 9 | 9 | 7 | 8 | 6 | 9 | 8 |
| You personally | 5 | 4 | 5 | 5 | 6 | 4 | 7 |
| The teachers | 4 | 1 | 5 | 4 | 5 | 2 | 3 |
| The Education Department | 1 | 7 | 3 | 7 | 1 | 6 | 2 |
| External Examination Bodies | 3 | 6 | 11 | 5 | 2 | 5 | 4 |
| Students | 8 | 8 | 8 | 11 | 9 | 7 | 9 |
| N | 18 | 19 | 11 | 18 | 19 | 14 | 99 |

(Note: Rank orders calculated from mean ratings of respondents.)

By contrast to the American study reported earlier (Schaffarzick, 1976: 4), control of "the ball game" as indicated by the three highest ranks is perceived to be in the hands of heads of department in all states, in the hands of teachers in Victoria and Tasmania, of principals in Queensland, South Australia and Tasmania, of deputy principals in Victoria and South Australia, and of the Education Department in New South Wales, Queensland and Western Australia; and it is perceived that authority is exercised over curricula by external examination bodies in New South Wales and Western Australia. Parents' and Citizens' Associations, individual parents, students, and the Curriculum Development Centre, (Canberra) are perceived as exercising least authority over curriculum decision-making in all states. On an Australia-wide basis, heads of department, the Education Departments and the teachers respectively occupy first, second, and third bases in terms of their control of the ball game called curriculum decision-making.

Given that there are differing perceptions from state to state of who controls the ball game in Australian secondary schools, what perceptions do the players themselves have of their roles in the game?

Table 3: Ranked Perceptions of Authority in Curriculum Decision-making by Respondent Category

| Authority | Respondents | | |
|---|---|---|---|
| | Principals | Department Heads | Teachers |
| Heads of department | | 1 | 1 |
| Deputy Principal | 6 | 7 | 5 |
| Principal | 1 | 3 | 2 |
| Individual parents | 10 | 11 | 11 |
| P & C Organization | 10 | 10 | 10 |
| Curriculum Development Centre, Canberra | 8 | 8 | 8 |
| You personally | 3 | 3 | 7 |
| The teachers | 1 | 2 | 3 |
| The Education Department | 5 | 6 | 4 |
| External Examination bodies | 4 | 3 | 5 |
| Students | 9 | 9 | 9 |
| N | 12 | 28 | 59 |

The respondents themselves differentially perceive authority in curriculum decision-making in the schools. Teachers consider that heads of department, the principal and teachers themselves have greatest authority; heads of department see themselves, teachers and principals in that order as having most authority, while principals consider that the principal and the teacher exercise equal authority in this regard. Principals perceive the authority of heads of department as much lower than do the other two groups of respondents.

Students, parents and the parent organization, and the CDC are perceived by all three categories of respondents as exercising the least authority over curriculum decision-making.

Concerning the relative emphases upon curriculum matters at general staff meetings, there is wide discrepancy in perceptions between principals and teachers (Table 4).

Table 4: Ranked Perceptions of Time Emphasis in Staff Meetings by Respondent Category

| Staff Meeting Topic | Respondents | | | |
|---|---|---|---|---|
| | Principals | Department Heads | Teachers | Total |
| Internal administrative matters | 2 | 1 | 1 | 1 |
| Administrative matters coming from external sources | 3 | 2 | 2 | 2 |
| Curriculum matters | 1 | 3 | 2 | 3 |
| N | 12 | 28 | 59 | 99 |

Whereas 92% of principals indicated that they perceived the time spent in general staff meetings is largely occupied by curriculum concerns, 93% of heads of department and 95% of teachers saw internal administrative matters as the major focus of such meetings.

Apart from perceptions of these general staff meetings, respondents were asked for information about meetings specifically convened to discuss curriculum. Although on a national basis it appears that principals, heads of department, and teachers all perceive that principals are the major initiators in convening curriculum meetings for the whole school staff, it appears that groups of teachers in Victoria take more initiative in convening such meetings than do the principals. (Tables 5 and 6)

Table 5: Ranked Perceptions of Conveners of Meetings to Discuss Curriculum by Respondent Category

| Meetings convened | Respondents | | | |
|---|---|---|---|---|
| | Principals | Department Heads | Teachers | Total |
| The principal | 1 | 1 | 1 | 1 |
| A staff committee | 2 | 2 | 2 | 2 |
| Individual teachers | 3 | 3 | 3 | 3 |
| Students | 5 | 5 | 5 | 5 |
| Persons external to the school | 4 | 4 | 4 | 4 |
| N | 12 | 28 | 59 | 99 |

Table 6: Ranked Perceptions of Conveners of Meetings to Discuss Curriculum by State

| Meetings Convened by | State | | | | | | ALL |
|---|---|---|---|---|---|---|---|
| | NSW (n=18) | VIC (n=19) | OUE (n=11) | SA (n=18) | WA (n=19) | TAS (n=14) | STATES (n=99) |
| The principal | 1 | 3 | 1 | 1 | 1 | 1 | 1 |
| A staff committee | 2 | 1 | 3 | 2 | 2 | 3 | 2 |
| Individual teachers | 3 | 2 | 2 | 3 | 3 | 2 | 3 |
| Students | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| Persons external to the school | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

## 2.0 Constraints on Curriculum Autonomy

Lack of finance is perceived by all three groups of respondents - principals, heads of department, and teachers - as the most severe problem in implementation of a planned program in Australian schools. The next most severe problems identified include "need to prepare for external/moderating exams" (principals), "lack of time for planning" (heads of departments and teachers), class size (teachers), and poor teachers and teaching (heads of department). Staff movement to other schools is perceived by all three groups as one of the most severe problems (Table 7).

Table 7: Perceptions of major problems in implementation of program

| | Respondents | | | | | |
|---|---|---|---|---|---|---|
| | Principals (n=12) | | Heads of department (n=28) | | Teachers (n=59) | |
| | % | Rank | % | Rank | $ | Rank |
| Need to prepare for external/moderating exams | 16.7 | 1 | 7.1 | 4 | 6.8 | 4 |
| Lack of finance | 16.7 | 1 | 17.9 | 1 | 15.3 | 1 |
| Lack of support staff | 8.3 | 3 | - | | - | |
| Staff movement to other schools | 8.3 | 3 | 7.1 | 4 | 6.8 | 4 |
| Lack of time for planning | 8.3 | 3 | 10.7 | 2 | 11.9 | 2 |
| Architecture or layout of the school | 8.3 | 3 | - | | 5.1 | 6 |
| Pressure from parents and community | 8.3 | 3 | - | | - | |
| Poor teachers and teaching | 8.3 | 3 | 10.7 | 2 | - | |
| Size of classes | - | | 7.1 | 4 | 11.9 | 2 |
| The school principal | - | | 7.1 | 4 | 5.1 | 6 |
| Architecture or layout of classrooms/teaching areas | - | | - | | 5.1 | 6 |

### 3.0 School Climate

One might hypothesise that, in those schools in which there are perceptions that communication patterns are good and where decision-making is subject to wide participation, curriculum development is likely to be promoted. In fact, this is the basis of the growing interest in "organizational development" as a major strategy for the improvement of school-based curriculum development (Bentzen et al, 1974; Cusick et al, 1976; Richardson, 1973; Steinhoff and Owens, 1976; Walden et al, 1975; Alderfer and Brown, 1975. Just how good is the current climate in secondary schools in terms of communication and participatory decision-making.

Evidence from the Curriculum Action Project data is suggestive of differing perceptions in these respects. Whereas principals perceive that decisions are participatory, heads of department agree less and teachers disagree with this perception. Whereas principals perceive that they accept and implement decisions, heads of department and teachers agree less. Whereas principals perceive that communication is good and meetings are not dominated by a few individuals, both heads of department and teachers disagree (Table 8).

Table 8: Perceptions about "School Climate"

| Aspect of climate | Principals | | Heads of department | | Teachers | |
|---|---|---|---|---|---|---|
| | Agree | Disagree | Agree | Disagree | Agree | Disagree |
| | f | f | f | f | f | f |
| 1. Participatory decision-making occurs (items 1, 2, 3, 7) | 34<br>(73.9%) | 12 | 59<br>(53.6%) | 51 | 103<br>(43.6%) | 133 |
| 2. Decisions accepted by principal and implemented (items 4, 8) | 18<br>(78.3%) | 5 | 37<br>(68.5%) | 17 | 68<br>(57.6%) | 50 |
| 3. Communication channels are good (items 5, 6) | 13<br>(56.5%) | 10 | 15<br>(37.5%) | 40 | 36<br>(31.0%) | 80 |

## SECTION IV: DISCUSSION OF FINDINGS

As already stressed, this is a preliminary report of pilot data of a relatively small sample, and therefore any findings at this stage must be regarded with great caution. Nevertheless, early data has highlighted certain aspects worthy of closer investigation.

(1) Although edicts concerning curriculum autonomy have reached educational headlines, there is evidence that the offering of autonomy may not have been matched by its acceptance. This varies somewhat between states. By a study of Table 2, one may contrast practices in NSW, Queensland and WA, with those in smaller states. One might even conjecture that increasing geographic distance from Head Office might have a significant influence upon communication procedures so as to reinforce the strength and frequency of head office documents!!

(2) In all states, heads of department are perceived as powerful authorities in curriculum decision-making. One might question the desirability of this situation in terms of the importance of planning the whole curriculum in school-based curriculum decision-making, rather than seeing the school curriculum segmented into subject territories lacking in cohesion and integration. (Case study data obtained so far strongly supports this expression of concern) This will be the subject of a subsequent Curriculum Action Project Research Report.

(3) How conducive are conditions in Australian secondary schools to promoting school-based curriculum decision-making?

  (a) Data suggests that general staff meetings do not provide the context for curriculum discussions and that the initiative for curriculum activities remains (with the exception of Victoria) in the hands of the principal. Staff development programs for strengthening participation in curriculum activities may be indicated as a need.

  (b) The most severe problems perceived by respondents in implementing school-based curriculum decision-making include lack of finance, lack of time for planning, and staff mobility. It behoves educational administrators to address themselves to the dilemma in which they have placed school personnel, by on the one hand decreeing school-based curriculum decision-making, and on the other, denying access to appropriate levels of support and staff continuity, which are basic to curriculum improvement.

(c) Perceptions of teachers concerning their lack of authority within curriculum decision-making is also reinforced by data from school climate items. Taken conjointly, these suggest that the prerequisite conditions for effective school-based curriculum decision-making, namely effective communication, participatory decision-making and decision implementation, are not being fulfilled. This highlights the importance of more widespread study of the strategies of "organizational development" as the bases for improvement of school-based curriculum decision-making.

## REFERENCES

Alderfer, C.P. and Brown, L.D. (1975) Learning from Changing: Organizational Diagnosis and Development, Beverly Hills, California: Sage Publications.

Bentzen, M.M. et al (1974) Changing Schools: The Magic Feather Principle, New York: McGraw-Hill

Curriculum Development Centre (1977) Triennial Program 1977-1979: New Directions in Curriculum Development with an Emphasis on Lower Secondary Schooling,Canberra, Australia: Curriculum Development Centre.

Cusick, P.A., Martin, W., and Palonsky, S. (1976) Organizational Structure and Student Behaviour in Secondary School, Journal of Curriculum Studies, 8, 1, 3-14.

Gross, N., Giacquinta, J.B., Bernstein, M. (1971) Implementing Organizational Innovations: A Sociological Analysis of Planned Educational Change, New York: Basic Books

Richardson, E. (1973) The Evolution of the Continuous Staff Conference, The Teacher, the School and the Task of Management, (E. Richardson), London: Heinemann, 312-33.

Schaffarzick, J. et al (1976) Current Issues, Problems and Concerns in Curriculum Development. A Report and Set of Recommendations Submitted to the National Council on Educational Research by the NIE Curriculum Development Task Force, Washington: The National Institute of Education, US Dept. of Health Education and Welfare.

Steinhoff, C.R. and Owens, R.G. (1976) Problems related to Techniques for Assessing Organization Development and Determining Intervention Style, Journal of Educational Administration 15, 2, 176-186.

Walden, J.C., Thomas, N.T., and Watkins, J.F. (1975) Organizational Climate Changes over Time, The Educational Forum, (Nov), 87-93.

## ACKNOWLEDGEMENTS

Without the generous cooperation of principals, heads of departments and teachers in 100 Australian secondary schools, this study would not have been possible.

The authors wish to acknowledge the invaluable assistance of the following colleagues:

Dr Kenneth Tye (Visiting Consultant from IDEA and UCLA, Semester 1, 1977) and his colleagues at IDEA, upon whose advice we drew freely;
Gail Thompson, formerly Senior Tutor in Education, Macquarie University;
Mick Bezzina, Research Assistant throughout the Curriculum Action Project;

Adrian Bradley, Research Assistant;
Dianna Williams, Research Assistant;
Shirley Colless, Research Typist;
Jean Platt, Research Typist.

We also wish to thank Mr Malcolm Rosier of the Durvey Section of ACER, who drew the sample of schools invited to participate;

The following people assisted in coding the data, and preparing questionnaires:

Stuart Harrison
Margaret Hercus
Mary Kelliher
Sean Mack
Roy Stanhope
Jo-Anne Watters
Helen Williams

# EVALUATION

# OF

# SCHOOL EXPERIENCE

# IN

# PRESERVICE TEACHER EDUCATION

John M. Atkinson, &
Geoffrey R. Trebilco
Rusden State College

Paper presented at the Annual Conference of the Australian
Association for Research in Education, Canberra, November, 1977

## ABSTRACT

Because school experience is such an important and expensive component of a teacher's pre-service preparation every effort should be made to ensure its maximum effectiveness. Evaluation has a crucial role to play in providing a sound basis for decisions about a school experience program.

The paper draws upon Stufflebeam's model of evaluation to provide a framework within which to describe and analyse approaches employed in evaluation of the school experience program at Rusden State College. Selected data are discussed in relation to decisions which they support. These examples have been chosen, where possible, with a view to their relevance beyond the particular institution.

# EVALUATION OF SCHOOL EXPERIENCE IN PRESERVICE TEACHER EDUCATION

John M. Atkinson and Geoffrey R. Trebilco
Rusden State College

## SCHOOL EXPERIENCE AND EVALUATION

The central place of school experience within preservice teacher education needs little substantiation. Quite aside from any theoretical rationale, ask any beginning teacher what was the most important, or most valuable, part of his professional preparation and the answer comes back almost invariably - "school experience" or "teaching practice".[1] Ask teachers or student teachers for suggestions to improve a preservice course and the most common proposal is to increase the proportion of time spent in school experience. (This is not peculiar to Australia - see, for example, Frankel, 1974.)

As well as being of such educational significance within teacher preparation, school experience in Australia also involves a financial operation of considerable proportions. In recent times it has become an increasing budgetary problem for teacher-training institutions. At Rusden State College, for example, estimated total costs for teaching practice[2] amount to over $500,000 for an education EFTS of 500 in the current calendar year. This is close to 9% of the total recurrent budget of the College. (In terms of course time teaching practice constitutes forty five days (or 7.5%) of a total of six hundred course days in a four year B.Ed. course.) School experience is a relatively expensive course component!

Furthermore, it represents an inviting target for the administrator's pruning shears, since a large proportion of the outlay is in off-campus payments, curtailment of which would not superficially appear to affect other operations of an institution.

It is little wonder that there is a mounting tendency in some (especially administrative) circles, both within tertiary institutions and outside, to question value for money insofar as school experience is concerned. There is certainly an obligation on those responsible for decision-making about school experience programs to endeavour to gain maximum effectiveness from those programs. Evaluation clearly has a crucial role to play in relation to school experience, especially if one accepts the conception of evaluation as "the collection and use of information to make decisions about an educational program." (Cronbach, 1963)

The present writers would argue that evaluation which is not geared to existing or prospective decision situations is of questionable value. Consequently, in presenting some of our own evaluation attempts relating to school experience, and considering some other possibilities, a decision-making rationale will be used.

A contribution to evaluation theory which has not been as widely recognised in this country as it deserves to be is that of Stufflebeam and the Phi Delta Kappa National Study Committee on Evaluation (1971). The all-embracing scope of the CIPP (Context, Input, Process, Product) model, as it has come to be known, makes it somewhat forbidding. It is, in many ways, perhaps most useful at macro levels of evaluation (e.g. system, or national). Nevertheless, it does have applicability on a small scale as well.

While the writers do not pretend that the somewhat piecemeal evaluation ventures outlined subsequently in this paper in any way adequately represent evaluation according to the Stufflebeam model, some of his conceptualisations do provide a valuable framework within which to consider them.

## THE STUFFLEBEAM FRAMEWORK

The decision-making basis of the CIPP model draws heavily on the work of Braybrooke and Lindblom (1963). Figure 1 summarises Stufflebeam's conceptualisation of four differentiated decision settings.

Decision-making in the metamorphic setting is rarely, if ever, possible in educational situations. It is certainly not relevant to the present paper. The other three settings are all applicable. Most decisions made in connection with a school experience program are homeostatic, aimed at maintaining normal balance. Changes in this setting take the form of minor remedial adjustments, such as taking steps to locate additional supervising teachers to cope with an increased enrolment of, for example, physical education students. Incremental decision settings occur wherever there is a concern for continuing development of a program through small modifications. Adjusting the placement of school experience within a one-year course, acting to improve liaison with supervising teachers and introducing

---

1. The terms "school experience", "teaching practice" and "the practicum" tend to be used interchangeably. In this paper "teaching practice" will be used to mean that time spent in schools under the guidance of a paid supervising teacher. "School experience" will be used more broadly to encompass all preservice course activities and experiences (including teaching practice) which take place in the schools. "Practicum" will not be used at all!
2. These take the form of payments to supervising teachers, College academic staff salaries apportioned to the school experience program, staff travelling reimbursements, secretarial, accounting and administrative services, library and curriculum materials support, printing, stationery, postage and telephone expenses, and servicing of College accommodation, plant and equipment used in connection with teaching practice.

| | | |
|---|---|---|
| HIGH<br>INFORMATION | **HOMEOSTASIS**<br>Activity: Restorative<br>Purpose: Maintenance<br>Basis: Technical standards and quality control | **METAMORPHISM**<br>Activity: Utopian<br>Purpose: Complete Change<br>Basis: Overarching theory |
| GRASP<br>LOW | **INCREMENTALISM**<br>Activity: Developmental<br>Purpose: Continuous Improvement<br>Basis: Expert judgment plus structured inquiry | **NEOMOBILISM**<br>Activity: Innovative<br>Purpose: Inventing, testing, and diffusing solutions to significant problems<br>Basis: Conceptualization, heuristic investigation, and structured inquiry |
| | SMALL | DEGREE OF CHANGE LARGE |

Figure 1: Decision-Making Settings (Stufflebeam et al., 1971, p.62)

an instrument for diagnosing and reporting student teaching weaknesses are decisions which might fall into this setting. Occasionally a decision for more fundamental change may be taken: this is neomobilistic decision-making. A change from a consecutive to a concurrent course, or a decision to substitute peer-group microteaching for a substantial proportion of school experience would be examples of this setting.

As with any arbitrary classification instances are likely to occur which lie somewhere between the idealised settings identified. Stufflebeam argues further that in order to develop an evaluation model which is appropriate to all educational decision-making situations, some way of categorising actual decisions (as distinct from decision settings) is needed. If this can be done, then it may be feasible to produce a generalisable evaluation design which can be applied to all decision situations within a given category. He proposes a classification of decisions employing two dichotomies: whether the decision pertains to ends or means, and whether it refers to intentions or actualities. This results in four decision types corresponding to which are proposed four types of evaluation.

Figure 2 illustrates this conceptualisation which has interesting parallels with the data matrix proposed by Stake (1967).

| | INTENDED | ACTUAL |
|---|---|---|
| ENDS | PLANNING DECISIONS<br>(To determine objectives)<br>supported by<br>CONTEXT EVALUATION | RECYCLING DECISIONS<br>(To judge and react to attainments)<br>supported by<br>PRODUCT EVALUATION |
| MEANS | STRUCTURING DECISIONS<br>(To design procedures)<br>supported by<br>INPUT EVALUATION | IMPLEMENTING DECISIONS<br>(To utilise, control and refine procedures)<br>supported by<br>PROCESS EVALUATION |

Figure 2: Decision Types & Related Evaluation (adapted from Stufflebeam et al., 1971)

## EVALUATION AND DECISION-MAKING IN THE RUSDEN SCHOOL EXPERIENCE PROGRAM

The school experience program at Rusden State College has been in a state of continuing development for the past six or seven years. From a (then) standard format of one three-week block round each term in a one-year education course it has changed into the program described in Appendix A, and is continuing to change. The changes have been effected following many decisions made in an incremental setting and a few which might be regarded as neomobilistic. Between changes and for aspects not altered by particular changes, homeostatic decision-making proceeds on a continuing basis.

This section of the paper will consider the types of information on which decisions have been based and how this information has been obtained; it will also discuss some selected examples of various decision types in relation to evaluation information which has influenced them. Some of this information has been collected deliberately and formally; some of it on the other hand has accumulated informally and, to some extent, unintentionally. There has clearly been other potential information which would have been useful had it been available.

The available information has been diverse both in kind and in the manner by which it has been acquired. In describing the evaluative processes by which information has been deliberately accumulated one could, at different times, reasonably invoke many of the contrasting terms which make up the dichotomies of current evaluation terminology (focus-portrayal, preordinate-responsive, goal directed-goal free, formative-summative, experimental-illuminative, etc.) which McGaw (1976) effectively summarised. The writers, being loath to commit themselves heavily to a polarised position at any time, believe this is as it should be.

## Context Evaluation

Stufflebeam (1971) sees context evaluation as incorporating both (i) comparison of intended and actual system performance and (ii) searching outside the immediate system for opportunities and pressures to engender improvement within it. As a result enlightened planning decisions can be made. A planning decision might take the form of modifying objectives (in the broad sense of ends to be pursued) or of deciding that no change to objectives is indicated. In either event, structuring, implementing and/or recycling decisions might ensue.

### Comparison of intended and actual system performance

(a) Student questionnaires which have been administered annually over a number of years have consistently included sections seeking information about perceived value of school experience and preferred arrangement of school experience within the total course structure, as well as inviting free comment. Much more detailed questionnaires used in 1975 and 1976 have been used to collect data from students in the concurrent B.Ed. program after their initial period of one day per week in schools in Year 3 of their course, and from both Year 4 B.Ed. students and Dip.Ed. students at the conclusion of their two day per week school experience.

In the case of Year 3 students information obtained has related to such aspects as nature of activities undertaken in the school and distribution of time between them, extent of guidance and assistance from both college and school supervisors, preferred arrangement of teaching practice, extent of satisfaction with school experience and reasons for wanting or not wanting to remain in the same school in the following (fourth) year.

Responses on both occasions have indicated that

- close to 40% of students have felt they needed more guidance in their initial period of teaching practice;
- third year students have been less favourable than fourth year students in their ratings of supervisors' helpfulness;
- the majority have not been observed by or had discussions with a College staff member during their nine days in schools during Term 3 of third year;
- almost 20% have rated their school experience as "hardly worthwhile" or in one or two cases "not worthwhile".

These results have not led to any change of objectives, but have led to some implementing decisions. Detailed written guidelines relating to the third year teaching practice segment have been given to students and supervisors in 1977. In addition, an effort has been made to strengthen support for the students by college staff both in classes and by visiting more in the schools. Consideration being given to more substantial change is referred to subsequently (p. 8).

Information gathered from Dip.Ed. and Year 4 B.Ed. students has included extent and perceived value of feedback on teaching performance from both college staff and school supervisors, rate of induction into teaching full lessons and nature of activities which preceded this, involvement in school activities other than scheduled class teaching, reasons for and reaction to remaining in the one school or changing schools, preferred teaching practice arrangement, and extent of satisfaction with school experience. Results relating to teaching practice arrangement are taken up on p.4 . One interesting set of findings relates to the "graduated experience" policy described in Appendix 1. Over 90% of students (in both courses) observed a lesson taken by the supervisor with the class concerned, and approximately 60% had some small-scale active involvement with the class before teaching their first full lesson. Nevertheless, the first full lesson was taken within two weeks (i.e. four days in the school) by about 80% of students. This represents a rather less "graduated" approach than has been envisaged by the College, but was no cause for concern since 90% of students reported that they would not have wanted their rate of initiation to be any different than it was, and most of the remainder would have preferred it to be more rapid than it was!

No action has been taken following these findings, since for a minority of students their introduction to "flying solo" was more gradual than for most and this, too, was as they generally would have wished it.

(b) There has been no regular written survey data collected from teachers in the schools who are involved in the school experience program, although two "special" studies are referred to subsequently. Various forms of organised meetings and seminars, involving supervising teachers and Rusden staff, held both at the College and in individual schools, have had an evaluative purpose as well as facilitating the running of the school experience program through enhanced communication.

(c) Informal discussion and/or observation provide an important avenue of context evaluation. While contact between supervising teachers, student teachers and College staff is common in any teacher education program, its program-shaping effect is probably heightened at Rusden by two factors. Firstly, as there are students in schools on three or four days of practically every week of the academic year (and some outside it), staff from the College are in and out of schools continually. Secondly, these same members of College staff are effectively the decision-makers regarding school experience organisation. Thus the conclusions which they come to regarding school experience, shaped by their on-going contacts with students and supervising teachers, interact with the more formally collected evaluative data in determining developmental modification in the program. The aspects about which formal data have been collected in recent student questionnaires have to a large measure been determined by informally collected observations and comments.

(d) Three special studies (in the sense of not being regular sources of information) have been carried out recently. The first of these was a survey in mid-1975 of a sample of 129 supervising teachers and teaching practice organisers from a randomly selected half of the schools taking Rusden students. It asked for a preference among five alternative teaching practice arrangements for the forty-five days of paid supervision effectively allowed by the funding system then operating. (This same choice had been given late in 1974 to a random sample of students completing their pre-service course, and was subsequently given to a further sample of students late in 1976.) Table I compares results (expressed as percentages) for these groups.

TABLE 1: Preferred Mode of Teaching Practice Organisation (percentages)

| TEACHING PRACTICE ARRANGEMENT | SUPER. TEACHERS (n=129) | STUDENT TEACHERS 1974 (n=91) | STUDENT TEACHERS 1976 (n=31) |
|---|---|---|---|
| 1. Two days per week in schools | 20 | 30 | 18 |
| 2. Three rounds of three weeks | 19 | 2 | 0 |
| 3. One term full time in school | 19 | 19 | 12 |
| 4. A three week round in Term 1, and two days per week in Terms 2 and 3 | 8 | 17 | 12 |
| 5. Two days per week in Terms 1 and 2 and a three week round in Term 3 | 34 | 32 | 58 |

Both supervising teachers and student teachers have shown a preference (stronger in the case of students) for a system incorporating two days per week in schools for a large proportion of the time. The traditional pattern of three week block rounds had relatively little support among supervisors and virtually none among students. There is, however, a strong indication (confirmed by meetings with supervising teachers held since the survey , and by responses to student questionnaire items calling for suggestions) that the most favoured system would comprise two days per week for the majority of the time, with block practice in the latter stages. This information has resulted in certain structuring decisions. Students are now being given the option of one or more full weeks of teaching practice in certain weeks when schools are operating but there are no College classes, in lieu of an equivalent number of days within the two days per week arrangement. There are difficulties associated with organising a block of two or more consecutive weeks, since this would have ramifications for the total timetable structure of the College. Alternatives which might lead to a more satisfactory decision are currently being evaluated.

The teacher survey also sought an effectiveness comparison between the two days per week arrangement of teaching practice and block teaching rounds, according to eleven different criteria. (Student teachers have not been asked to make the same comparison since whereas most supervising teachers would have had some experience of each system, this is not the case for the students.) Responses to this part of the questionnaire, shown in Table 2, may explain the support for a combination of the two modes.

In interpreting Table 2, note should be taken of a tendency on the part of some respondents (21 of the sample of 129) to "block vote" for one mode or the other on every criterion. While these split evenly between the two modes (11 for two days per week and 10 for block rounds) they appear to indicate a disinclination to consider each criterion separately. The effect of this is a reduction in the apparent perceived superiority of one mode or the other on a particular criterion.

Despite this, it can be said that

(i) the two days per week system was seen to be more suitable for promoting relationships with supervising teachers, staff generally and pupils, and to a lesser extent for involvement in school activities and development of confidence;

TABLE 2: Comparison by Supervising Teachers of Two Modes of Teaching Practice Organisation (percentages)

| CRITERION | TWO DAYS PER WEEK MORE SUITABLE | NO DIFFERENCE BETWEEN THE TWO SYSTEMS | BLOCK PRACTICE MORE SUITABLE |
|---|---|---|---|
| 1. Realistic preparation for teaching | 48 | 15 | 37 |
| 2. Development of effective relationship between student teacher and supervisor | 60 | 21 | 19 |
| 3. Development of productive relationships with school staff generally | 57 | 18 | 25 |
| 4. Development of productive relationships with pupils | 67 | 10 | 23 |
| 5. Involvement in overall activities of the school | 48 | 21 | 31 |
| 6. Acquisition of Teaching Skills | 40 | 40 | 20 |
| 7. Development of the flexibility to be effective in a variety of types of school | 11 | 20 | 69 |
| 8. Ability to apply theory to practice | 39 | 44 | 17 |
| 9. Knowledge of the non-classroom aspects of a teacher's duties | 34 | 28 | 38 |
| 10. Opportunity to be responsible for teaching units of work | 26 | 15 | 59 |
| 11. Development of confidence | 50 | 34 | 16 |

(ii) little difference was seen between the two modes in terms of realistic preparation for teaching, acquisition of teaching skills and ability to apply theory to practice, although two days per week was slightly favoured on all these counts;

(iii) no appreciable difference was seen in provision of knowledge of non-classroom duties;

(iv) the two day per week system was seen to be inferior in terms of developing flexibility to be effective in a range of school types and in provision of opportunity for teaching units of work.

These results were generally predictable and reflect the issues considered when Rusden opted for the two day per week system. The "flexibility" problem, associated with the placement of most students in the same school throughout the year, was recognised, but considered to be outweighed by other positive factors. Steps have been taken to increase students' opportunity to broaden their range of experience of schools through other course activities. Efforts have also been made to inform supervisors of ways which have been found effective for enabling students to teach units of work.

The second "special" study was carried out late in 1976. It was an attempt to elicit from student teachers, supervising teachers and teachers who had completed their education studies at Rusden two years previously (hereinafter referred to as "former students") their views as to the effectiveness of the Rusden school experience program. One section of the questionnaire used listed a number of outcomes (see Appendix 2) to which it might be considered pre-service school experience should contribute. The outcomes were generated through consultation with staff, student teachers and supervisors, and from other relevant sources including results of a survey of principals carried out by the Victorian High School Principals Association. Respondents were asked to rate the potential outcomes according to two criteria:

(a) the extent to which each outcome could be achieved through pre-service school experience limited to 45 days;

(b) the extent to which each outcome was being achieved in the Rusden school experience program.

Both ratings were based on a 3 point scale ( 1 = very little, 2 = partly, 3 = substantially). Table 3 shows the mean ratings of each group of respondents for both criteria.

Some interpretations of the results shown in Table 3 follow:

(i) Not surprisingly, the overall rating of actual achievement was generally lower than perceived potential achievement. However, if a rating of 2.0 or greater is taken as a reasonable criterion of success of the program to achieve a given outcome, then the current students believed 90% of the objectives were potentially achievable at or beyond the criterion level but that only 45% were being achieved to that level.

The corresponding figures for supervisors were 60% and 45%, and for former students, 15% and 25%.

(ii) The perceived discrepancies between potential and actual achievement tended to be much greater for former students (mean discrepancy = .39) and current students (mean discrepancy = .32) than for supervising teachers

TABLE 3: Mean Ratings of Twenty Potential Outcomes of School Experience by Student Teachers, Supervising Teachers and Former Students

| STATEMENT NUMBER (refer Appendix 2) | EXTENT TO WHICH OUTCOME IS ACHIEVABLE | | | EXTENT TO WHICH OUTCOME IS BEING ACHIEVED | | |
|---|---|---|---|---|---|---|
| | Students (n=72) | Supervisors (n=80) | Former Students (n=46) | Students | Supervisors | Former Students |
| 1 | 2.51 | 2.26 | 2.17 | 2.28 | 2.08 | 2.00 |
| 2 | 2.74 | 2.65 | 2.57 | 2.44 | 2.44 | 2.37 |
| 3 | 2.61 | 2.42 | 2.39 | 2.21 | 2.42 | 2.02 |
| 4 | 2.38 | 2.52 | 2.43 | 1.90 | 2.63 | 2.16 |
| 5 | 2.14 | 1.98 | 1.78 | 1.69 | 1.82 | 1.42 |
| 6 | 2.40 | 2.28 | 2.20 | 1.77 | 2.18 | 1.73 |
| 7 | 2.13 | 1.79 | 1.96 | 1.62 | 1.60 | 1.30 |
| 8 | 1.99 | 1.72 | 1.93 | 1.58 | 1.62 | 1.47 |
| 9 | 2.32 | 2.20 | 2.32 | 1.91 | 2.00 | 1.68 |
| 10 | 2.31 | 2.07 | 2.15 | 1.97 | 2.00 | 1.65 |
| 11 | 2.06 | 1.59 | 1.89 | 1.76 | 1.37 | 1.44 |
| 12 | 2.30 | 1.83 | 2.00 | 2.03 | 1.66 | 1.66 |
| 13 | 2.40 | 2.18 | 2.46 | 2.28 | 1.99 | 1.96 |
| 14 | 2.13 | 2.01 | 2.02 | 1.97 | 1.88 | 1.50 |
| 15 | 2.33 | 2.02 | 1.93 | 2.03 | 1.96 | 1.70 |
| 16 | 2.50 | 2.33 | 2.24 | 2.18 | 2.44 | 1.87 |
| 17 | 1.65 | 1.38 | 1.52 | 1.56 | 1.35 | 1.24 |
| 18 | 2.12 | 1.57 | 1.91 | 1.85 | 1.33 | 1.57 |
| 19 | 2.34 | 2.07 | 2.40 | 2.14 | 1.95 | 1.89 |
| 20 | 2.47 | 2.39 | 2.22 | 2.31 | 2.26 | 2.14 |

(mean discrepancy = .12). Among possible explanations for this is that the involvement of supervisors in providing the program may contribute to their more optimistic view of what is achieved compared with what could be achieved. It did appear from replies to written response sections of the questionnaire that some former students might not have effectively differentiated school experience from the pre-service course as a whole. Some differences between former and present students' responses may reflect the effect of incremental change to the program. For example, the higher achievement perceived by present students in regard to outcome numbers 11 and 13 may indicate that opportunities for learning about alternative patterns of school organisation and for teaching units of work are now provided more effectively (refer p. 5 ).

In general, those outcomes for which greatest discrepancy between potential and actual achievement is perceived could be regarded as areas where ways of effecting improvement most need exploration. Number 7 ("flexibility to be effective in a variety of school situations") crops up again here (cf. p. 5 ). Others include the elusive Number 9 ("relating of theory to practice"), Number 6 ("development of productive relationships with pupils...") and Number 8 ("involvement in extra-classroom activities..."). It would be interesting to have comparative results relating to other school experience programs, since the two days per week arrangement over an extended period would seem more conducive to achievement of Numbers 6 and 8 then would, say, traditional three-week blocks.

An interesting item is Number 4 ("development of effective relationships between student teacher and supervisors"). Whereas supervisors' results somewhat anomolously indicate actual achievement to exceed potential (perhaps the item was treatening), student teachers' results on this item showed the second highest shortfall between potential and actual achievement recorded for any of the outcomes. This is somewhat surprising as direct questions about helpfulness of supervisors have always been answered favourably in the past.

Another point of interest is that the general outcome, Number 20 ("effective overall preparation for a teaching career") comes closest of any item to achieving perceived correspondence of potentiality and actuality when considered across the three groups.. There seems to be a greater tendency to perceive specific inadequacies in the program than general inadequacy.

(iii) When the twenty potential outcomes were ranked on achievability:

- Nos. 2,3,4,16 and 20 were in the top eight for all three groups;
- the same eight items - Nos. 5,7,8,11,12,14,17,18 - were ranked lowest, but in varying orders, by all groups;
- all but the above eight items were accorded mean achievability levels above the arbitrarily-chosen standard of 2.0 by all groups;
- Nos. 8 and 17 were accorded mean achievability levels below 2.0 by all groups.

Spearman-rank order correlation coefficients between the orderings on the criterion of achievability were as follows:

Supervising teachers - Student teachers: $r = 0.92$
Supervising teachers - Former students: $r = 0.88$
Former students - Student teachers: $r = 0.81$

The high values for these coefficients suggest that the orderings have credibility and might be used to provide some guidelines in making planning decisions about objectives for school experience.

The third "special" study, results of which were not yet available, involved the development of an instrument which sought to identify student teachers' perceptions of various aspects of their course, of schools and teaching, and of themselves in relation to these items. Students were asked to respond by means of a Likert scale to 46 statements. The statements included some adapted from the Minnesota Teacher Attitude Inventory (1951), some from the Teaching Anxiety Questionnaire developed by Ball (1976) and some invented for the purpose. Responses were collected from as many as possible of the students entering Year 3 of the B.Ed. course prior to the start of the 1976 academic year, from the same population again just before the commencement of their teaching practice, and again the following year when school experience was complete. A factor analysis of responses on the first administration has provided a basis for monitoring changes in perceptions exhibited by various identifiable groups of students at different stages of their professional preparation. Such information could clearly have a bearing on planning decisions relating to school experience. Somewhat related bases for context (or product) evaluation with regard to school experience are provided by the use of self concept measures (Coulter, 1974) and various anxiety and attitude scales (e.g. Poole and Gaudry, 1974; Ball, 1976).

## Opportunities and Pressures from Outside the System

External pressures are manifold with respect to teacher education in general and school experience programs in particular. Pressures in the form of expressed or implied expectations regarding teacher preparation come from politicians, the media, organised groups and individuals. The current emphasis on basic skills is an example of this.

Alternatively, external pressures may take the form of constraints imposed by agencies or conditions at varying distances from the system. For Rusden's school experience program these have included, in recent years:

- an undersupply in available supervising teachers with the result that competence has been virtually inapplicable as a criterion of selection for the task. (This situation is likely to change as numbers of students in training are reduced, and with the altered payment conditions resulting from the recent academic salary award - although industrial unrest provoked by the latter has resulted in some supervisors withdrawing their services.)
- funding levels and/or supervisor payment rates which assume a maximum forty five days supervised teaching practice, allied with Secondary Teacher Registration Board requirements which stipulate a minimum of forty five days supervised teaching practice for registration.
- the need to spread the student teacher load on schools reasonably evenly over the year.
- the need, within a concurrent program, to achieve an acceptable organisation of school experience vis-a-vis both academic and education course units.
- shortage of teaching accommodation at the College with a resulting pressure to distribute school experience over the week in order to even out the demands on the space available.

Pressures have a tendency to manifest themselves (although it is usually advantageous to anticipate them as soon as possible); opportunities, on the other hand, usually have to be sought. External ideas for improving or enriching the school experience program have come from such sources as study visits (both outward and inward), attending conferences and reading. Opportunities quite often arise at a school for a particularly valuable activity or experience. Changes to constraints may present opportunities: for example, changes to the system of supervisor payment have opened up the prospect of utilising highly effective supervisors more intensively than was possible when Teachers' Tribunal regulations governed the maximum sum payable per year.

It is clearly important that both pressures and opportunities are effectively perceived and analysed.

## Input, Process and Product Evaluation

Whereas context evaluation is regarded by Stufflebeam (1971) as general and systematic, the other three identified evaluation types are perceived as operating only after a planning decision has been made to introduce some change. In general, the more drastic the change and the more restricted the knowledge of how the change might best be accomplished, the more comprehensive and formal is the evaluation required.

Input evaluation is the gathering of information needed for designing a strategy to accomplish new or changed objectives. The following is a simple example of this.

In the B.Ed. course students commence supervised teaching practice at the rate of one day per week in the third term of Year 3. To phase in the concurrent course structure in 1975, schools had to contend with a considerable overload of student teachers in Term 3 of that year. Prior to the decision to implement concurrency being taken, schools associated with Rusden were circularised in 1974 to see if they supported the concurrency concept and would be prepared to accept the overload in Term 3 of the transition year. The favourable response obtained was a modest but crucial piece of information resulting from input evaluation and enabling a relatively major structuring decision to be taken.

The process of identifying and evaluating alternative ways to achieve a block of teaching practice with minimal disruption to other programs (referred to on p.4) is another example of input evaluation.

Process evaluation is carried out to detect faults in a designed strategy as it is being implemented and to maintain a record of the procedure as it occurs. Some of the student questionnaire data described earlier (for example, the proportions of time spent on various types of activity within the school) constitutes process evaluation when collected to monitor a modification to the program. Much process evaluation currently occurs informally.

Product evaluation has the purpose of ascertaining what has been achieved, either at the end of a cycle or during it, following a change made within the system. The distinction between product evaluation and context evaluation thus made is mainly on the basis of scope and timing: once a change becomes a standard feature of the program the evaluation of its effects is encompassed by context evaluation. In practice it may often be extremely difficult to isolate the outcomes of a specific change, particularly in an incremental setting. Some of the data collected after the initial experience of one day per week teaching practice for Year 3 students, reported on p.3, might properly be categorised as product evaluation. Evidence of a lower level of satisfaction with teaching practice, compared with Year 4 students, supported by opinions expressed by supervising teachers, is information which has led to a re-examination of this segment of the program. Should the implementing decisions which have been made in an effort to overcome the perceived deficiencies (refer p.3), prove unsuccessful, a recycling decision may then be made to terminate this form of practice in favour of some alternative strategy.

## CONCLUSION

The writers have sought to describe and analyse the ways in which evaluation has been used as a basis for decision-making concerning the school experience program at Rusden State College. Some of the basic conceptualisations of the evaluation theory proposed by Stufflebeam et al. (1971) have been employed to provide a framework for this examination. It may be appropriate to conclude with some brief comments about the efficacy of this framework.

Stufflebeam and his co-authors are the first to concede that the models and strategies they have put forward are imperfect and incomplete. There seems little cause to dispute, however, that both the models and the detailed strategies that have been developed (but not explored in this paper) represent a significant contribution to evaluation theory. Their broad conception of the underpinning role of evaluation in rational decision-making provides in itself a basis for guiding evaluation initiatives in useful directions. Cognisance of the distinctions between various decision settings and decision types further refines this process and helps to increase the probability that information needed for effective decision-making is available at the right time.

The distinctions between evaluation types are also useful, but tend to become blurred in practice. Frequently it may be convenient (and even necessary) to collect two or more of context, input, process and product data at the one time and even by the one instrument. It may sometimes be difficult to decide whether it is context evaluation, or one of the others which is being carried out at a particular time. This does not matter: what is important is that decisions of consequence be made on the basis of the best information that can be provided.

## APPENDIX 1

## RUSDEN STATE COLLEGE

## A Short Description of the
## SCHOOL EXPERIENCE PROGRAM

The College runs two types of teacher preparation program for intending secondary teachers. For students in the four year Bachelor of Education course, education units and school experience are spread over Years 3 and 4 in a concurrent structure. In addition, a one-year Diploma in Education course is offered to graduates of other institutions.

Content of the two programs is basically similar, albeit differently organised, but the concurrent structure permits attempts at integration of "academic" and "education" work as well as providing the option of taking the equivalent of up to one and one-third years' study in education. (The minimum education component in that course is equivalent of one year's full-time study, as for the Dip.Ed.)

In both programs school experience comprises forty-five days supervised teaching practice, essentially on a two days per week basis, together with a range of supplementary experiences.

Consistent with a policy of graduated experience, students spend some six to twelve days in a wide variety of schools for directed observation and discussion before starting supervised practice. A College staff member normally accompanies each group of students who may visit two or more high schools (metropolitan and near-country), a primary school, a kindergarten, a technical school, and perhaps an independent secondary school and/or a special school for periods of from half a day to four of five days. These visits are scheduled so as to relate to an integrated program of introductory education studies, 'Teaching and Learning', undertaken at the College during this period. This program provides further (vicarious) experience through use of videotaped classroom material as well as incorporating some peer-group micro-teaching and/or other means of practising teaching and planning skills. Microteaching in the school situation is provided for many students.

Formal teaching practice is organised slightly differently in the B.Ed. concurrent program than for Dip.Ed. For the B.Ed. students, teaching practice commences on a one day per week basis in Term 3 of third year and continues at the rate of two days per week during Terms 1 and 2 of fourth year. In the Dip.Ed. program, teaching practice on a two days per week basis starts in the fourth week and continues until four weeks from the end of the course.

Many students remain in the same school throughout the teaching practice phase of both programs, although changes of school can and do occur for a variety of reasons (including the student's desire to experience a different type of school). Most students are placed in metropolitan high schools, some in technical schools, and a small proportion in independent schools (although steps are usually taken to ensure that the latter group spend some time in a State school). Some students choose country schools and travel to and from the city each week.

Consistent with the graduated experience policy, supervising teachers are encouraged to arrange for students to commence with observation of lessons and progress to limited responsibility (e.g. supervision of individual or group work, teaching of part lessons, team teaching, etc.), until ready to take full lessons. A gradual increase in teaching load and planning autonomy over the year is recommended.

Further school or related experience is afforded through a proportion of the wide range of elective units which comprise about one third of each student's education studies. This additional experience ranges through such activities as the teaching of computer programming, development and implementation of a short-term remedial program, work with migrant children, participation in open classroom programs, development and trialling of an environmental studies unit, or work in community recreation programs including drama and physical education.

Students in the B.Ed. program, in particular, have the opportunity to undertake individual or group projects which can provide still more experience in schools or other educational settings.

## APPENDIX 2

## LIST OF POTENTIAL OUTCOMES USED IN SCHOOL EXPERIENCE PROGRAM EVALUATION

1. The gaining of knowledge of secondary school structure and organisation.
2. Effective preparation for lessons the student is to teach.
3. The development of basic teaching skills.
4. The development of effective relationships between the student teacher and his/her supervising teachers.
5. The development of productive relationships with school staff generally.
6. The development of productive relationships with pupils in classes taken by the student teacher.
7. The development of flexibility to be effective in a variety of school situations.
8. Involvement in extra-classroom activities of the school.
9. The relating of theory to practice.
10. The gaining of proficiency in class management including "discipline".
11. The understanding of alternative patterns of school organisation.
12. The development of sensitivity to and techniques for providing for individual differences.
13. The gaining of experience in planning and teaching complete units of the curriculum.
14. The acquisition of a broad repertoire of teaching methods or techniques.
15. The development of a sense of professional responsibility.
16. The development of confidence as a teacher.
17. Development as a participating staff member able to contribute constructively in such areas as curriculum planning, policy formulation, etc.
18. An understanding of the relationship between the school and its societal context, e.g. pupil background, career prospects, parent and community expectations.
19. Development of an ability to capitalise on available resources both from within the school and from without.
20. Effective overall preparation for a teaching career.

## REFERENCES

Ball, I.L. The Efficacy of Human Relations Training in Modifying Student Teacher Anxieties. Department of Psychology, State College of Victoria - Rusden. Mimeographed research report, 1976.

Beeson, G.W. Student teachers, college lecturers and the schools: a new approach to teaching practice. Council of Teachers College Staff Associations (Victoria). Mimeographed report of symposium, Melbourne, 1973.

Braybrooke, D. and Lindblom, C.E. A Strategy of Decision. New York: The Free Press, 1963.

Cook, W.W., Leeds, C.H. and Callis, R. Minnesota Teacher Attitudes Inventory. New York: The Psychological Corporation, 1951.

Coulter, F. The effects of practice teaching on professional self image. The Australian Journal of Education, 1974, 18, 149-159.

Cronbach, L.J. Course Improvement Through Evaluation. Teachers College Record, 1963, 64, 672-683.

Frankel, E. Evaluation of Lehman College undergraduate education program by teacher alumni. Bronx, New York: City University of New York, 1974.

McGaw, B. Evaluation of practice as atheoretical research. Paper read at the annual conference of the Australian Association for Research in Education, Brisbane, November, 1976.

Poole, C. and Gaudry, E. Some effects of teaching practice. The Australian Journal of Education. 1974, 18, 255-263.

Stake, R.E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O. and Provus, M.M. Educational Evaluation and Decision Making. Itasca, Ill.: Peacock, 1971.

Worthen, B.R. and Sanders, J.R. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.

# HOW CHANGE IS MEASURED, AND WHY IT USUALLY ISN'T:
# IMPLICATIONS FOR EDUCATIONAL RESEARCH.

Glenn Rowley

La Trobe University

## ABSTRACT

It is a truism that the purpose of education is to bring about changes in students. One might imagine, then, that the measurement of change would occupy a central place in evaluative research, but this has never been the case. Measurement of change has come to be regarded as an area so fraught with difficulties that it is perhaps best avoided altogether. This paper examines and evaluates these difficulties, and recommends, where appropriate, suitable procedures to deal with them. The major problems identified are

(1) the unreliability of measures of change;
(2) the problem of metric; and
(3) correlational problems and regression effects.

These problems are explained and illustrated with emphasis on the difficulties they create for the conduct of empirical evaluative research.

We live in a world of change, and our task as educators is to help in preparing children to cope with their changing world. The task has not become any easier in recent years, and there is no reason to think that it will in the near future. In 1970, Alvin Toffler wrote that

> Western society, for the past 300 years, has been caught up in a fire storm of change. This storm far from abating, now appears to be gathering force. Change sweeps through the highly industrialized countries with waves of ever accelerating speed and unprecedented impact (p. 9).

In reacting to these changes, our schools themselves have changed (although some will say not enough). New curricula, new teaching methods and new organizational structures have all been responses to the changing demands on the schools. Education is changing. We educate children about change, so that they can cope with change; perhaps even so that they can become the agents of change.

Cliches, of course, they all are. But they serve to illustrate the point that education is for change, it is about change, and, in a very fundamental sense, education _is_ change. When we educate a person, we change that person. Gagné (1977) defines learning as

> ...a change in human disposition or capability, which persists over a period of time, and which is not simply ascribable to processes of growth (p. 3).

In this sense, one's effectiveness as a teacher is a direct function of the nature, number, and size of the changes one brings about in one's pupils. Teaching methods and curricula will be evaluated by examining the changes which result from their implementation. Educational research will be an activity which seeks to detect, measure, and account for changes in the pupils who are the centre of the educational process. Measurement of change will occupy a central place in educational evaluation and empirical research.

Well, the argument may sound plausible enough, but the conclusions don't seem to fit very well with what we see when we observe the activities of educational evaluators and researchers today. The actual measurement of change is carefully avoided, and, in large part, the science of experimental design has become, in fact, the art of making inferences about change without actually measuring it. Why should this be so? What is it about the measurement of change which makes competent researchers avoid it so studiously? The purpose of this paper is to provide some answers to these questions.

The factor which stands out above all others is that the measurement of change has proven to be exceedingly difficult and full of traps for the unwary. The ease with which one may be misled can be illustrated by a simple example which originated with Lord (1956), in an article which stimulated a great deal of the attention which has been paid to the subject in the twenty years since. Lord considered a hypothetical case in which a group of students were tested with parallel forms of a test just before and just after their summer vacation. The intention was to compare the pre-test and post-test scores of individual pupils in order to determine the amount of change, or growth, which had taken place in the intervening time period: Figure 1 is a representation of the scatterplot which resulted.

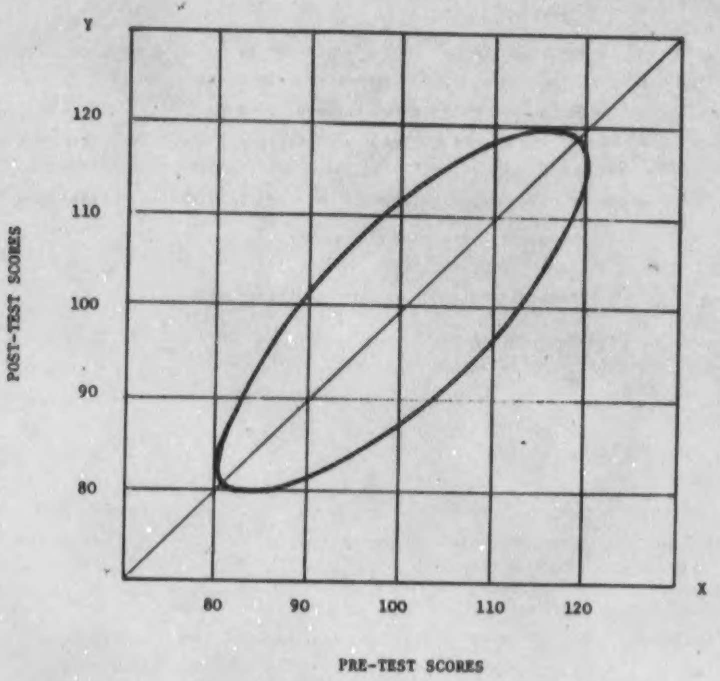


Figure 1. Hypothetical relationship between pre- and post-test achievement scores. (Adapted from Lord, 1956)

Upon examining Figure 1, we note that the lowest-scoring pupils on the pre-test (those who scored 81 or less) all improved their score upon subsequent re-testing, while those who scored highest on the pre-test (i.e. 119 and over) all declined in score. With a little thought, we might conclude that a strange phenomenon had been occurring during the summer - most of those who initially scored below average improved their scores, while the opposite happened for those scoring above average on the pre-test. Computation of change scores (where change score = final score - initial score) for each pupil would confirm this fact.

Thus far, we have taken no account of measurement error which unfortunately is present in all tests. If, instead of waiting for the summer vacation to pass, we had immediately re-tested the pupils with the parallel test, the scatterplot would have had the same general appearance as that shown in Figure 1, and for all we know, it may have been identical. (Since the distributions of scores are the same on the two occasions, this would be the case if the correlation represented by the scatterplot was the same as the test-retest reliability.) If this, in fact, proved to be the case, what should we conclude? The scatterplot is precisely the one we would expect if no pupils had changed at all, so that common sense would suggest that we ignore all the individual change scores, and conclude that the individual gains and losses we have observed are merely the results of measurement error. As Lord commented, this "conclusion might conceivably be wrong in the case of a single exceptional individual, but it is nevertheless the only general conclusion justified by the data available". (Lord, 1956, p. 423).

Where does this leave us? In this particular case, it seems that we would do better to ignore the individual change scores, and base our conclusions on the characteristics of the group. But how many of us would have done this? And what if there were no group testing, but just two scores obtained from one individual? The potential to be misled by computed change scores is very great indeed.[1]

In sections to follow, we will examine three of the major problems facing investigators who wish to measure and make inferences about change. Although solutions to these problems will be suggested where appropriate, detailed consideration, and recommendation of appropriate procedures cannot be made in a paper of this length. Instead, they will be included in a separate paper (currently in preparation).

The arguments which follow involve some statistical reasoning, but statistical notation has been kept to a minimum in order to make the issues comprehensible to non-statisticians. To the extent that it is necessary to include formulae, the notation which has been adopted is essentially the same as that used by O'Connor in his (1972) review article. Every observed score is regarded as being made up of a true (error-free) component, and a random error component. True scores are indicated by capitals, observed scores and error scores by lower-case letters. Scores of interest include the following:

| Type of score | Observed score | True score | Error score |
|---|---|---|---|
| Initial status | $x$ | $X$ | $e_x$ |
| Final status | $y$ | $Y$ | $e_y$ |
| Gain score | $g$ | $G$ | $e_g$ |

The gain score $g$ is defined as the difference between final and initial status. Two gain scores are of interest; the observed gain score,

$$g = y - x, \tag{1}$$

and the true gain score,

$$G = Y - X \tag{2}$$

Our problem, as in any measurement situation, is to make inferences about true scores when the data at our disposal consist only of observed scores. The usual relationships among true scores, observed scores, and error scores are assumed, so that

$$x = X + e_x \tag{3}$$

and

$$y = Y + e_y \tag{4}$$

from which it follows that

$$g = G + e_g \tag{5}$$

---

[1] The phenomenon at work in this example is statistical regression, and is one which creates particular problems in the interpretation of change scores. It is a subject to which we shall return in a later section.

## MAJOR PROBLEMS.

### 1. Reliability.

Gain scores are notoriously unreliable (Bereiter, 1963; Lord, 1963; Linn and Slinde, 1977). Whenever we subtract one measurement from another, as has long been recognized in the physical sciences, the potential error becomes larger, while the magnitude of the quantity measured is typically much smaller. The resulting measurement, then, contains a much larger proportion of error than either of the two from which it was derived.

Reliability has traditionally been defined in operational terms as the correlation between two parallel tests, or (more generally), between two independent measures of the same quantity. Suppose then, that, in addition to measures $x$ and $y$, we have parallel measures $x'$ and $y'$. The gain scores of interest are

$$g = y - x,$$

and

$$g' = y' - x'.$$

The reliability of these gain scores is given by the correlation between $g$ and $g'$, and may be shown (McNemar, 1958) to be given by the formula

$$r_{gg} = \frac{S_y^2 \; r_{yy} - 2S_x \; S_y r_{xy} + S_x^2 \; r_{xx}}{S_y^2 - 2S_x \; S_y \; r_{xy} + S_x^2}, \qquad (6)$$

where $S_x^2$ and $S_y^2$ are the variances of $x$ and $y$ respectively, $r_{xx}$ and $r_{yy}$ their respective reliabilities, and $r_{xy}$ is the correlation between $x$ and $y$.

Formula (6) expresses the reliability of the gain scores in terms of observable quantities: the reliabilities of the pre- and post-scores $x$ and $y$, their intercorrelation, and their variances. It is the appropriate formula to use in computing the reliability of gain scores, and is cited frequently (e.g. Linn and Slinde, 1977; Stanley, 1971, p. 387).

In formula (6) we may note that the numerator and denominator differ only by the inclusion of the reliability terms in their numerator. It follows that if the pre- and post-scores $x$ and $y$ had perfect reliability, so too would the gain scores (unless, of course, the denominator was zero, which could only happen if the two sets of scores were perfectly correlated - in this unlikely event, the reliability would be indeterminate).

Many texts (e.g. Brown, 1976; Lemke and Wiersma, 1976; Mehrens and Lehmann, 1973; Thorndike and Hagen, 1977) do not cite formula (6) for the reliability of gain scores but provide instead a simplification of it which is based on the assumption that $x$ and $y$ have equal variances:

$$r_{gg} = \frac{r_{xx} + r_{yy} - 2r_{xy}}{2\,(1 - r_{xy})} \qquad (7)$$

Formula (7) is certainly simpler to use than formula (6), and it is in some ways more instructive, but it is not appropriate if the assumption of equal variances is not met, and frequently this will be the case. For example, if the scores were obtained before and after some treatment or training programme, the effect of which had been to increase the lower scores to a greater extent than the higher scores, there would be a smaller variance in the post-test than in the pre-test and formula (1) would be inappropriate. This situation could arise through the application of a mastery learning strategy (for which a reduction in achievement variance is an avowed aim), or through the use of a test upon which the highest scores on the pre-test were approaching the maximum score (the "ceiling effect"). The point is not that the variances would never be equal, but that they frequently will not be, and that the assumption is quite unnecessary. The data needed for the use of the more accurate formula (6) would be available to the researcher, and should be used.

The effect of unreliability of $x$ and $y$ on the gain score reliability is quite dramatic. Consider for example, a case in which we had moderately reliable measures ($r_{xx} = r_{yy} = 0.8$), with a moderately high intercorrelation ($r_{xy} = 0.7$), and suppose that, in this instance, the pre-test and post-test variances were indeed the same. The reliability of the gain scores could then be comput[illegible]ing equation (9), to yield

$$r_{gg} = \frac{0.8 + 0.8 - 2 \times 0.7}{2(1 - 0.7)}$$

$$= 0.33.$$

This is a fairly typical result. At least with respect to their own reliability, the measures used were probably perfectly adequate. The gain scores which they yield measure up to none of the normally accepted standards of reliability.

One might imagine that the obvious way to increase the reliability of $g$ would be to increase the reliabilities of the tests $x$ and $y$. This is true, but only to a limited extent. Suppose that we improve the tests until their reliabilities are 0.9 - probably approaching the limits of feasibility for most types of tests. The problem is that, other things being equal, one effect of this will be to increase the correlation between $x$ and $y$, and from equation (7) it can be shown that the effect of increasing $r_{xy}$ is to decrease the reliability of the gain scores. In the above case, we could expect $r_{xy}$ to increase by a factor of (0.9/0.8) to take the value 0.7875. (This is a straightforward application of the correction for attenuation - see, e.g., Nunnally, 1967, p. 219). The reliability of the gain scores would therefore become

$$r_{gg} = \frac{0.9 + 0.9 - 2 \times 0.7875}{2(1 - 0.7875)}$$

$$= 0.53$$

In practical terms, this is probably about as much improvement as one could reasonably expect to obtain just by improving the reliabilities of $x$ and $y$. To obtain further improvements we would need to increase these reliabilities without increasing the correlation between $x$ and $y$. Indeed the highest gain-score reliabilities would come about if we were able to increase (or at least maintain) $r_{xx}$ and $r_{yy}$, and progressively reduce the inter-correlation $r_{xy}$. Methods of item-selection advocated for the development of measures of change would have this effect (Bereiter, 1963; Saupe, 1966). Of course, in this case we would have serious doubts that the test was measuring the same thing on the two occasions, and therefore whether it was even appropriate to regard the difference scores as measuring "change" in anything. Bereiter (1963) refers to this problem as the unreliability - invalidity dilemma: the more reliable the change scores become, the greater our doubts about their validity. It is a dilemma which admits of no entirely satisfactory solution.

An interesting consequence of this dilemma has been pointed out recently by Overall and Woodward (1975) who argue that the unreliability should not be regarded as a problem in studies designed to measure change induced as a result of a treatment. They consider the case in which measures are taken before and after treatment, and the difference between the two means is then tested for significance. The higher the correlation between pre-test and post-test scores, the lower the reliability of the difference scores, but the greater the likelihood that significant change will be detected. In the limit, as they point out, the power of the significance test is a maximum when the reliability of the change scores is zero!

Although the "paradox" proves, on close examination, to be less of a paradox than it appears, it does, at least serve to illustrate a most important fact about change scores. While their unreliability may make them unsuitable for making inferences about (and comparisons between) individuals, it does not necessarily limit their usefulness in situations where one wants to make inferences about group means. The reliability of a sum (and, therefore, of a mean score) is naturally much greater than that of the scores from which it is derived. (This is why single items, each of which, taken by itself, may be quite unreliable, can be combined together to form a test which yields quite reliable scores.) Many psychological measures (particularly measures of personality) are of low reliability, but are validly used for making group comparisons. The same is true of measures of change. As Bereiter (1963) commented, "it is necessary to distinguish between the meaningfulness of average changes and the meaningfulness of individual differences in change. In (examples cited) there would undoubtedly be large average changes in test scores that would be unambiguous in meaning . . . It is the individual deviations from the mean trend that are meaningless, or as we shall prefer to say, of no interest" (p. 12).

The unreliability of change scores must, then, be a matter of concern in research which aims at identifying and making inferences about individuals; tasks for which change scores are manifestly unsuitable. It is also a problem in any investigation of the correlates of change, since all correlations will be substantially reduced by the unreliability. It need not be regarded as such a serious problem in studies of group changes, since in these studies it is the reliability of the group mean which is of crucial importance. In evaluating curricula, to the extent that we call upon empirical evidence, we are more likely to seek evidence about group changes than about individual changes, and to this extent at least, change scores may be regarded as adequate.

## 2. The Problem of Metric.

If we are to subtract one score from another, and interpret any difference which occurs as evidence of "change", atleast two assumptions appear to be involved. The first assumption is that the scores measure the same attribute on the two occasions. In the case of physical measurements (height, weight, temperature, etc.), this assumption may not be too difficult to accept, for these are areas in which operational definitions are available, and not seriously questioned. With educational and psychological measures, this is far from the case. An educational or psychological test provides an example of indirect measurement, in which we typically observe a student's behaviour under a certain set of stimuli, and make inferences about the amount of some attribute, skill, or ability that the student possesses. The responses themselves are usually not the attribute of interest, but the evidence upon which we base our estimates of it. We may seriously question whether an achievement test measures the same attribute on the second administration as it does on the first, particularly if the content of the test is closely related to the objectives of the teaching which intervened. For example, an item which required very high-level cognitive skills before instruction may be a matter of mere recall afterwards. If students come from a variety of previous backgrounds (as in Grade 7 at a high school), a pre-test may reflect, more than anything else, their previous schooling and opportunity to learn, while the same test used as a post-test, following instruction, would reflect, one hopes to a large extent, their abilities and efforts during the intervening period. The researcher who wishes to infer "change" on the basis of successive test scores must first establish reasons to be confident that the test was measuring the same thing on the two occasions, and this is a tricky problem indeed.

A related but not necessarily distinct, problem is the "practice effect" - the effect of the pre-test itself. Stanley and Hopkins (1972, pp. 138-139) summarize studies in which gains averaging up to 7 IQ points have been recorded on intelligence tests merely as a result of retesting with the same or a parallel form of the test. It would be easy to mistakenly infer gains in achievement on the basis of increased test scores, when all that has happened is that the students have become a little more skilled at taking the tests. A further problem is that gains due to practice appear to be greatest for students who are less experienced at taking tests (Knapp, 1960) so that a naive researcher might even detect systematic relationships among difference scores even when no change at all has occurred. In summarizing the literature on practice effects, Stanley and Hopkins (1972) arrived at the following conclusions:

1. Practice effects are more pronounced with persons of limited educational background or experience.
2. The effects are greater on speeded tests.
3. The effects are greater on a repeated test than on a parallel form of the test. There is little or no practice effect after the second retest.
4. The greater the time interval between tests, the smaller the effect. There appears to be little practice effect for an interval greater than three months.
5. Other things being equal, the effects appear to be slightly greater for examinees of high mental ability.
6. For a group of typical examinees, the average practice effect is usually $.2\sigma$ or less in magnitude (p. 139).

Inferences about change made on the basis of test scores would be strengthened if the researcher took at least the following precautions:

1. Work with students who are reasonably experienced at taking tests.
2. Use relatively unspeeded tests.
3. Use parallel forms of a test, rather than retesting with an identical test.
4. Allow an appropriate time interval (preferably three months or more between testings).

Campbell and Stanley (1966) discuss the effect of practice as a confounding factor in experimental design. Other factors in the Campbell and Stanley scheme which might lead a researcher to wrongly infer change as a result of treatment include history and maturation (real changes which had nothing to do with the treatment), instrumentation (changes in the measuring instrument itself and in what it measures), and statistical regression (where the group has been selected on the basis of extreme scores on the pretest). Experimental designs which control for these factors are provided by Campbell and Stanley (1966) whose work is summarized in most educational research texts (e.g. Ary, Jacobs and Razavieh, 1972; Best, 1977; Borg and Gall, 1971; Kerlinger, 1973; Lehmann and Mehrens, 1971).

The second assumption involved in measuring change by the differences between successive scores is that the scale of measurement has at least interval properties, i.e., that a score difference of, say, 10 points at one part of the scale describes the same difference in amounts of the attribute measured as a score difference of 10 points at another point on the scale. If we believe this not to be so, we will in fact be allocating equal change scores to changes we believe not to be equal, and this is scarcely a rational procedure. The problem is of course, that there are very few educational measures (and probably no test scores) for which the assumption could be strongly defended. For some test scores, it is quite clearly invalid, e.g. percentiles. For others, it may appear reasonable (e.g. age or grade norms).

Until recently, little attention was paid to this problem. Lord, who has contributed so much to thinking on the measurement of change, regarded it more as a technical problem which the test constructor should overcome - "the experimenter should build his test in such a way that he is willing to treat numerically equal gains in different parts of the score scale as if they were at least approximately equal" (Lord, 1958, p. 450). He recognized the arbitrary nature of the assumption involved, but took the view that it was up to the investigator to either assume or demonstrate such equality, or to avoid making such comparisons:

> Useful comparisons of the gains of students who start out at quite different parts of the score scale require either an arbitrary assumption or an empirical demonstration that numerically equal intervals in different parts of the score scale are actually "equal" in some meaningful and useful sense. The reader will have to supply his own arbitrary assumption or empirical demonstration for each set of data; or else he will have to forego comparisons between the gains of individuals who start with quite different initial scores (Lord, 1956, p. 435).

The logic of the advice is not in question, but it does seem to gloss over the difficulties involved in creating such a scale. Until recently, this writer was inclined towards the view that scores expressed in grade-equivalents or age-equivalents were amongst those for which the assumption was most defensible, and perhaps the only ones for which it was truly valid. After all, a year's growth is a year's growth, and we have related scores from a raw-score metric about which we know very little to age and grade levels, which are indisputably interval in character. Surely the resulting scores would have interval properties!

Well, it seems that the answer is "no". Linn and Slinde (1977) found, for a wide variety of standardized achievement tests, that the standard deviations of grade-equivalent scores were approximately twice as great for children at Grade 6 level as they were for children at Grade 2 level. What are the implications of this? One is that "the naive expectation of a gain of one grade-equivalent limit in a year's time" (p. 140) is quite unjustified. To maintain the same relative positions in relation to their peers, students of differing ability would have to progress at quite different rates, as measured in grade-equivalents. A student at the 80th percentile, in order to remain at that (relative) level, would have to "gain" considerably more than one grade-equivalent each year. A student at the 20th percentile could maintain that relative position by gaining somewhat less than one grade-equivalent per year. The only students who could gain exactly 1.0 grade-equivalents each year, and maintain the same position relative to their peers are those who begin (and remain) at the 50th percentile.

Suppose then, that we have two students, one whose initial performance is at the 80th percentile and another at the 20th, and that during the course of an experiment each increase their grade-equivalent scores by 1.0. The former has gained rather less than would normally be expected of him in a year, while the latter has gained considerably more. Does it make sense to assume that the two performances represent equal "gains"? Probably not. But what else can we do? One suggestion considered by Linn and Slinde (1977) is "vertical equating" of tests - the scores are standardized in such a way as to make the standard deviations the same at each grade level. This too leads to problems. In particular, it means that the grade-equivalent score corresponding to a chance-level performance (i.e. random responding) undergoes quite substantial increases from year to year. In the most dramatic example shown, a student who responded randomly to the arithmetic section of the Metropolitan Achievement tests in Grades 5 and 6 would increase his grade-equivalent score by almost 2.0 in that time when there has quite obviously been no improvement in performance (Linn and Slinde, 1977, Figure 3).

It appears, then, that the interval properties of grade- and age-equivalent scores are extremely doubtful, so that measures of change based on them would be of rather dubious validity. It is not at all clear for what scores the assumption of interval properties could be justified, and, in general, it may be concluded that the problem of metric is a rather more serious one than has previously been recognized.

### 3. Correlational problems and regression effects.

Frequently, investigators are interested in identifying correlates of change - characteristics of the individual which are associated with change on a curricular outcome such as academic achievement, and which therefore might be useful in identifying the types of students most likely to profit from a particular treatment or educational program. To know, for example, that students with a particular interest and ability profile are the greatest gainers (in terms of test scores, or interest measures) from a year's instruction in PSSC Physics would seem to be a particularly valuable piece of information to possess; in the short term, for use in counselling students, but in the long term, for use in the development and modification of further curricular materials.

Unfortunately, studies which have attempted to identify the correlates of change scores have not been very successful in this regard. There are many reasons for this. One has already been discussed - the fact that change scores are in general extremely unreliable means that they do not (and cannot) correlate highly with any variable. This is problem enough in itself, but is only the beginning of the difficulties we must face. In fact, an experienced investigator could usually predict quite successfully before data were even collected which variable would be most closely related to change - the students' initial scores on the criterion measure!

There are two reasons why this should be so. One is that there is present in many tests what may be described as a "ceiling effect". If a student's score on a test is determined (as it normally is) by the number of correct answers given, then students who obtain high scores on the initial testing have very little opportunity to improve their scores on the post-test. If there are large gains to be made, they can only be made by the students who gain low scores on the pre-test. Students with high pre-test scores cannot achieve high change scores, no matter how much they may learn or otherwise benefit from the instruction. For this reason, among others, gain scores are frequently found to correlate negatively with initial standing. It is possible to dismiss this as merely a "technical difficulty", which can be remedied by the use of a more suitable test, but this would be making light of a serious problem. If a test has the level of difficulty necessary to measure accurately as a pre-test, then it very probably will exhibit a "ceiling effect" when used as a post-test. If its level of difficulty makes it suitable for use as a post-test, it is likely to be a very difficult pre-test, and therefore unreliable because of random guessing. The problem can only be satisfactorily overcome by using different tests at each stage, each at the appropriate level of difficulty. But if we do this, how can we then make inferences about changes?

There is a second reason why gain scores tend to be negatively correlated with initial standing, and this applies even when the test measures adequately on both occasions. The gain scores and the pre-test scores share the same errors of measurement, even though in opposite directions. Thus, the observed correlation between gain and initial score is a biased estimate of the correlation between true gain and true initial score. In fact, there is a simple relationship between the observed-score and true-score covariances, which can easily be demonstrated to be as follows (see, e.g. O'Connor, 1972):

$$C(x, g) = C(X, G) - S^2_{e(x)} \qquad (8)$$

where $S_{e(x)}$ is the standard error of measurement of $x$, and can easily be estimated from the reliability of $x$.

We see, then, that the covariance between observed gain and observed initial score is systematically less than the covariance between the corresponding true scores, and would be negative even when the true-score covariance is zero! The same, of course, is true of correlation coefficients, so that findings of zero or negative correlations between gain and initial score are really quite meaningless. This result has long been known, and was first demonstrated by E.L. Thorndike (1924). The problem may be overcome in two ways: measuring the error variance, and adding that term back in (Zieve, 1940); or using parallel pre-tests, one of which is used to measure initial status, and the other to compute the gain scores (Thorndike, 1966). The first procedure makes a statistical adjustment, while the second designs the enquiry in such a way as to avoid the problem of shared errors. An investigator who wishes to look at the relationship between change and initial standing needs to adopt one or other of these strategies; otherwise his conclusions will be very much biased in the direction of finding negative relationships. (It should be pointed out though, that sharing of common errors of measurement only occurs when measures of change are correlated with initial status, and is not a problem when change is correlated with other variables which are measured independently of the change scores.)

Up to this point, we have assumed the presence of measurement error. Not all of our problems, however, arise from this source. Lord (1956, 1963) has demonstrated clearly that the tendency for gain to be negatively correlated with initial standing is present, even in the absence of test inadequacy or measurement error. Figure 2 (adapted from Lord, 1963) illustrates this point. Weights of individuals in successive months are plotted, and we may assume that the weights are measured without error or at least with negligible error. In addition, we note that measures of weight have interval properties, so that there is no problem with units of measurement, and that there is no conceivable "ceiling effect". Just the situation, one might surmise, in which we could expect to be mercifully free of measurement problems!

In Figure 2, initial weight ($x$) is plotted horizontally, and final weight ($y$) is plotted vertically. The diagonal line through the origin (labelled $g = 0$) is in fact the line $y = x$, and points falling on this line correspond to people who have neither gained nor lost weight in the interval between weighings. The diagonal lines above this ($g = 5$, $g = 10$, $g = 15$) correspond to gains of 5, 10 and 15 kilograms respectively, while those below ($g = -5$, $g = -10$, $g = -15$) correspond to losses of 5, 10 and 15 kilograms. It is important to note that there has been no overall gain or loss by the group, taken as a whole, although there have been some (possibly random) changes in the individuals making up the group.

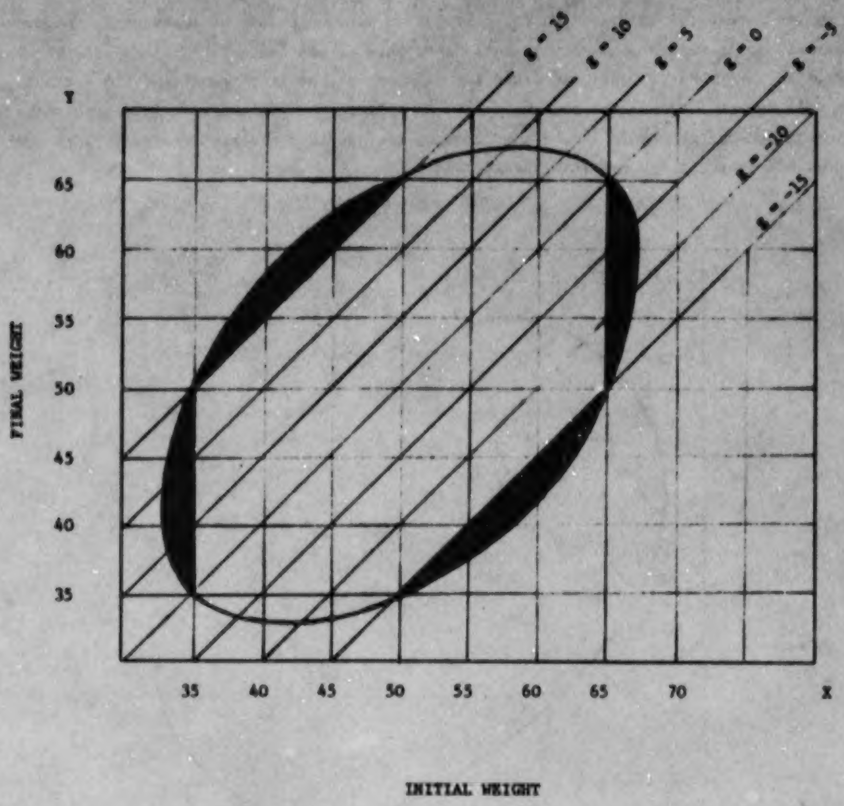


Figure 2. Hypothetical scatterplot showing changes in weight.
(Adapted from Lord, 1963)

Given the above, one might imagine that there would be very little that was systematic about the changes that took place. If so, one's imaginings would be far from the truth! Let us note the following points:

1. All individuals with initial weights less than 35 kilograms were gainers (the shaded area at lower left).
2. All individuals with initial weights greater than 65 kilograms were losers (the shaded area at upper right).
3. All individuals who gained over 15 kilograms were below average weight initially (the shaded area at upper left).
4. All individuals who lost over 15 kilograms were above average weight initially (the shaded area at lower right).

These four facts illustrate clearly that, even in a situation in which measurement error is absent, and in which there are no overall (systematic) changes, there is a definite, negative relationship between measured change, and initial standing. The phenomenon is a common one, and is, familiar enough to statisticians as "the regression effect". However, as Lord (1956) remarked, "(to) attach a name to the observed phenomenon does not change the facts of the case" (p. 422). The facts of the case are that initial score correlates negatively with gain in weight and that this correlation will be reflected in the correlations of gain in weight with any other characteristics which are correlated with initial score. For this reason, in asking research questions about the correlates of change, it is advisable in most cases to "control" in some way for initial score. Methods of doing this have been discussed in detail by Cronbach and Furby (1970), Lord (1963), Manning and DuBois (1962), O'Connor (1972), and Tucker, Damarin and Messick (1966).

Basically, these methods involve the researcher in the use of what have become known as "residual change scores", partial correlations, and (possibly) analysis of covariance. Many may feel that this unnecessarily complicates what really ought to be a relatively simple problem. The following example illustrates the ease with which one may be misled by simple correlations with raw change scores.

Suppose we are interested in examining the correlation between amount of learning (as measured by change in level of achievement) and some independently measured attribute of the learner. For the sake of argument let the attribute of interest be described as "aptitude", and let us suppose that it correlates quite strongly with initial score, as we might well expect. Since, for reasons outlined earlier, change itself is likely to share a negative correlation with initial score, we should expect that change, too, would correlate negatively with aptitude. Figure 3 shows a possible shape of the resulting scatterplot.

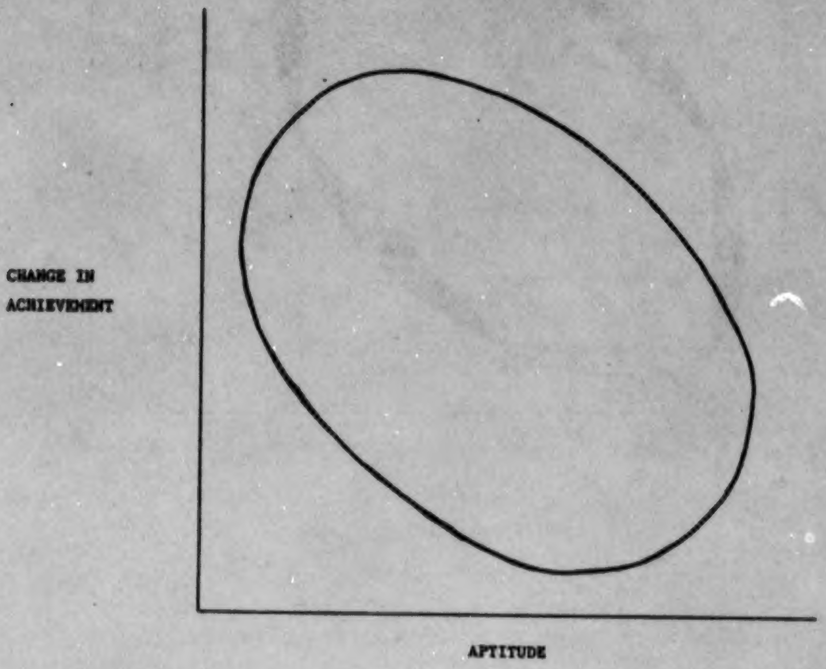


Figure 3. Hypothetical scatterplot showing the relationship between aptitude and change in level of achievement (no account being taken of initial score).

Clearly, if this is as far as our analysis went, we would note the negative correlation, conclude (perhaps with some puzzlement) that the most able learned least, and the least able learned most and (probably) look for a more profitable line of research. This would be a mistake.

Since aptitude is positively correlated with initial score, we will find that the high aptitude pupils mostly had high initial scores, and the low aptitude pupils mostly had low initial scores. Controlling for initial score may be equated to examining the relationship between aptitude and change for groups of children who have the same (or similar) initial scores. Suppose we break the group up into five sub-groups, each one relatively homogeneous with respect to initial score. Fig.4 illustrates a possible result. We notice that the overall correlation remains negative (as exemplified by the overall downward slope), but that within every sub-group, the correlation between aptitude and change is positive and substantial. The same data appear to be leading us to two contradictory conclusions. Which one is correct?

This phenomenon has been described by a number of writers, notably Lord (1958, 1970), and the answer seems to be that both are correct, but that the latter interpretation is probably more meaningful. In general, a correlation coefficient is more easily interpreted if more outside variables are controlled. For example, suppose a researcher is investigating the correlation between height and musical ability, and that, in every real sense, there is none. If it happens that, on average, males have more height, but less musical ability than females (as well they might), then a combined group of males and females could exhibit a marked negative correlation, whereas within either group, no correlation exists (see Figure 5). The relationship between height and musical ability would be better described by the zero

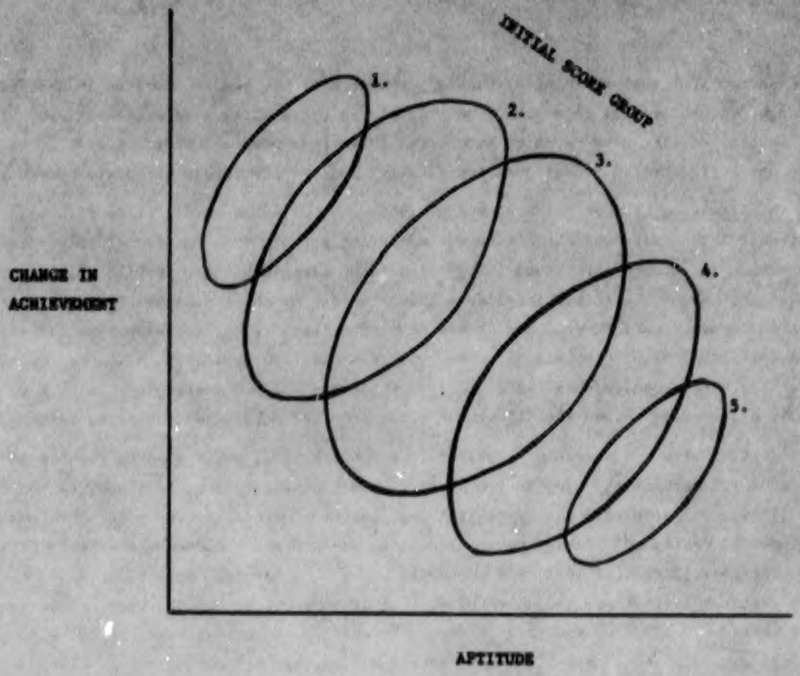


Figure 4. Hypothetical scatterplot showing the relationship between aptitude and change in level of achievement (controlling for initial score).

relationship (with sex held constant) than by the negative relationship, which is due entirely to sex differences. In the same way, the correlation between aptitude and change is probably more illuminating to us when initial score is held constant (i.e. using a partial correlation) then when initial score is ignored. The correlates of change are best investigated by the use of partial correlations, and any researcher considering entering the area would be well advised to make a thorough study of (at least) Lord (1958, 1963) and Cronbach and Furby (1970) before embarking.

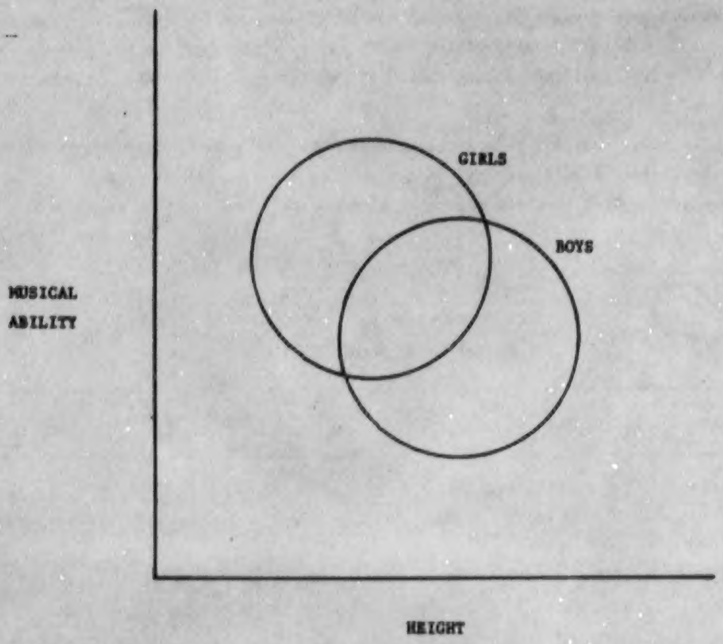


Figure 5. Hypothetical scatterplot showing the relationship between height and musical ability, for boys and girls.

## A CASE STUDY OF ERROR.

As might be expected from the preceding discussion, researchers who become involved in analysing and manipulating measures of change frequently fall into one or another of the traps which have been outlined. It would be neither profitable nor tactful to list studies which have erred in various ways. Instead, I will focus on a single, widely-publicized study which illustrates (albeit unintentionally) most of the errors which are commonly made.

Pygmalion in the Classroom, (Rosenthal & Jacobson, 1968) was published amid a blaze of publicity and favourable press comment. It described a research project conducted by the authors in a San Francisco elementary school from 1964 until 1966. Teachers in six grades were given information about certain of their pupils which was calculated to induce favourable expectations about the future intellectual development of those pupils. Over the next two years, all children in the classrooms involved were tested, and it was concluded that "(when) teachers expected that certain children would show greater intellectual development, those children did show greater intellectual development" (p. 83). In fact the specified children had been randomly selected from their grades, and therefore ought not to have differed systematically from their classmates, unless the difference arose from the artificially-induced teacher expectations!

Following the initial highly favourable reception accorded its findings, Pygmalion then began to attract attention on the grounds of the methodology used in the study. Some scathing reviews followed, two of the more forthright coming from Snow (1969), who concluded that Pygmalion had "performed a disservice to teachers and schools, to users and developers of mental tests, and perhaps worst of all, to parents and children whose newly gained expectations may not prove quite so self-fulfilling" (p. 199), and Thorndike (1968) who was even more blunt in declaring that the study was "so defective technically that one can only regret that it ever got beyond the eyes of the original investigators!" (p. 708). The critics focussed on two main issues: the quality of the data used in the study, and the opportunistic style of the data analysis. Pygmalion soon became one of the most severely criticized and staunchly defended pieces of educational research ever conducted.

Surprisingly, an aspect of the study which has attracted very little attention is the extent to which the conclusions of the study rested upon the analysis of raw gain scores. In fact, raw gain scores (the differences between test scores before and after the expectations were induced) constituted the dependent variables in the study. How then, did Rosenthal and Jacobson deal with the areas of difficulty described earlier in this paper? Let us examine them, one by one.

Unreliability. The only mention of reliability in the text occurs on pages 149 and 150, where the authors point out (quite correctly) that unreliability would make the detection of significant effects less, rather than more likely, and could not be used, therefore, to explain away statistically significant results. In a footnote, on page 163, we are told that the "over-all mean reliabilities" were .75 for verbal and total IQs, and .49 for reasoning IQs. However when "reliabilities" are finally reported (in the thirtieth of 31 tables tucked away in an appendix) we learn that the figures given are, in fact, correlations between pre-test and post-test one year later (i.e. $r_{xy}$ instead of $r_{xx}$ or $r_{yy}$, in the notation of pages 4 and 5). The reliabilities which really matter - those of the raw gain scores which are the dependent variables in the study - are simply not reported, and the reader is not provided with the information needed to compute them.

Metric. Throughout the text, only raw gain scores are reported and commented on. On some occasions, the experimental and control groups differed significantly; on others they did not. However, when we look beyond the text, to tables A1 to A9 in the Appendix, we find that the ranges of class means to be as shown in Table 1.

TABLE 1

RANGE OF CLASS MEANS FOR IQ SCORES
(Rosenthal & Jacobson, 1968)

| | Pre-tests | | | | Post-tests | | | |
|---|---|---|---|---|---|---|---|---|
| | Control | | Experimental | | Control | | Experimental | |
| | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| Low | 30.79 | 25.63 | 53.50 | 4.50 | 87.71 | 9.26 | 88.00 | 14.45 |
| High | 1[illegible]23 | 16.63 | 149.40 | 22.79 | 137.94 | 18.71 | 163.40 | 55.61 |

The foolishness of even reporting a class mean IQ of 30.79, with a standard deviation of 25.63, has already been noted (Thorndike, 1969). But to numerically equate "gains" from this level with "gains" made by classes who started, on average, over 100 points higher does seem to be placing rather more faith in the interval properties of IQ scores than most of us would be willing to do. Yet all of this is concealed in the text, where only gain scores are reported, and nowhere is it noted that such huge pre-test differences might make the gain scores uninterpretable.

*Correlations*. As a "check" on a hypothesis, the authors computed the correlations between gain and initial score. To their apparent surprise, the correlations were largely negative (p. 150). Table A31, which is intriguingly titled "Intellectual performance gain after one year as a function of pretest IQ", reports 60 correlation coefficients, 41 of which are negative. Of the 28 coefficients described as statistically significant, 26 are negative. The tendency for the correlations to be negative is reported as an interesting empirical finding, whereas in fact, it is nothing but a statistical artefact, and should have been recognized as such. In addition, correlations with various other measures were reported, ranging from classroom behaviours such as "curious", "interesting", "happy" and "affectionate" (p. 200) to the possession of Mexican facial characteristics (p. 127). No attempt was made to control for pre-test differences by using partial correlations, and what the reported correlations might mean is anybody's guess.

## CONCLUSIONS.

At least in our conclusions, we can be brief. Measuring change with educational or psychological tests is an exceedingly tricky business, and many researchers who have attempted it have ended up with egg on their faces. It can be done, and reasonable inferences can be made about change, but the use of common-sense methods of analysis probably will not be sufficient. The investigator whose fundamental research questions necessitate the use of change measures should not be deterred, but should make a careful study of (at least) the articles by Lord (1956, 1963), Cronbach & Furby (1970), O'Connor (1972), and Linn & Slinde (1977), and heed the advice contained therein.

## REFERENCES

ARY, D., JACOBS, L.C. & RAZAVIEH, A. *Introduction to Research in Education*, New York: Holt, Rinehart & Winston, 1972.

BEREITER, C. Some persisting dilemmas in the measurement of change. In C.W. Harris (ed.) *Problems in measuring change*, Madison: University of Wisconsin Press, 1963.

BEST, J.W. *Research in Education*, (3rd. ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1977.

BORG, W.R. & GALL, M.D. *Educational Research: an introduction* (2nd. ed.) New York: McKay, 1971

BROWN, F.G. *Principles of Educational and Psychological Testing* (2nd. ed.) New York: Holt, Rinehart & Winston, 1976.

CAMPBELL, D.T. & STANLEY, J.C. *Experimental and quasi-experimental designs for research*, Chicago: Rand-McNally, 1966.

CRONBACH, L.J. & FURBY, L. How should we measure "change" - or should we?. *Psychological Bulletin*, 1970, *74*, 68-80. See also Errata, *ibid*, 1970, *74*, 218.

GAGNÉ, R. *The Conditions of Learning* (3rd. ed.) New York: Holt, Rinehart & Winston, 1977.

KERLINGER, F.N. *Foundations of Behavioral Research*, New York: Holt, Rinehart & Winston, 1973.

KNAPP, R.R. The effects of time limits on the intelligence test performance of Mexican and American subjects. *Journal of Educational Psychology*, 1960, *51*, 14-20.

LEHMANN, I.J. & MEHRENS, W.A. *Educational Research: readings in focus*, New York: Holt, Rinehart & Winston, 1971.

LEMKE, E. & WIERSMA, W. *Principles of Psychological Measurement*, Chicago: Rand-McNally, 1976.

LINN, R.L. & SLINDE, J.A. The determination of the significance of change between pre- and post-testing periods. *Review of Educational Research*, 1977, *47*, 121-150.

LORD, F.M. Problems arising from the unreliability of the measuring instrument. In P.H. DuBois & G.D. Mayo (eds.) *Research Strategies for Evaluating Training*, AERA, Monograph Series on Curriculum Evaluation, Chicago: Rand-McNally, 1970, 79-93.

LORD, F.M. Elementary models for measuring change. In C.W. Harris (ed.) *Problems in Measuring Change*, Madison: University of Wisconsin Press, 1963.

LORD, F.M. Further problems in the measurement of growth. *Educational and Psychological Measurement*, 1958, *18*, 437-451.

LORD, F.M. The measurement of growth. *Educational and Psychological Measurement*, 1956, *16*, 421-437. See also Errata, *Ibid*, 1957, *17*, 452.

MANNING, W.H. & DU BOIS, P.H. Correlational methods in research on human subjects. *Perceptual and Motor Skills*, 1962, *15*, 287-321.

McNEMAR, Q. On growth measurement. *Educational and Psychological Measurement*, 1958, *18*, 47-55.

MEHRENS, W.A. & LEHMANN, I.J. *Measurement and Evaluation in Education and Psychology*, New York: Holt, Rinehart & Winston, 1973.

NUNNALLY, J.C. *Psychometric Theory*, New York: McGraw-Hill, 1967.

O'CONNOR, E.F. Extending classical test theory to the measurement of change. *Review of Educational Research*, 1972, *42*, 73-97.

OVERALL, J.E. & WOODWARD, J.A. Unreliability of difference scores: a paradox for measurement of change. *Psychological Bulletin*, 1975, *82*, 85-86.

ROSENTHAL, R. & JACOBSON, L. *Pygmalion in the Classroom*, New York: Holt, Rinehart & Winston, 1968.

SAUPE, J.L. Selecting items to measure change. *Journal of Educational Measurement*, 1966, *3*, 223-228.

SNOW, R.E. Unfinished Pygmalion. *Contemporary Psychology*, 1969, *14*, 197-199.

STANLEY, J.C. Reliability. In R.L. Thorndike (ed.) *Educational Measurement* (2nd. ed.) Washington: American Council on Education, 1971.

STANLEY, J.C. & HOPKINS, K.D. *Educational and Psychological Measurement and Evaluation*, Englewood Cliffs, N.J.: Prentice-Hall, 1972.

THORNDIKE, E.L. The influence of chance imperfections of measures on the relationship of initial score to gain or loss. *Journal of Experimental Psychology*, 1924, *7*, 225-232.

THORNDIKE, R.L. Review of Rosenthal & Jacobson's *Pygmalion in the Classroom*. *American Educational Research Journal*, 1968, *5*, 708-711.

THORNDIKE, R.L. Intellectual status and intellectual growth. *Journal of Educational Psychology*, 1966, *57*, 121-127.

THORNDIKE, R.L. & HAGEN, E.P. *Measurement and Evaluation in Psychology and Education* (4th. ed.) New York: Wiley, 1977.

TOFFLER, A. *Future Shock*, New York: Bantam Books, 1970.

TUCKER, L.R., DAMARIN, F. & MESSICK, S. A base-free measure of change. *Psychometrika*, 1966, *31*, 457-473.

ZIEVE, L. Note on the correlation of initial scores with gain. *Journal of Educational Psychology*, 1940, *31*, 391-394.

# SOME PROBLEMS OF CURRICULUM EVALUATION AT GRADE 12 LEVEL

R.G. ROWLANDS
The Victorian Universities and Schools Examinations Board

## Introduction

In an invited submission to the Victorian Education Department's Curriculum Services Enquiry, Professor B. McGaw has outlined the problems which he perceives in the provision of curriculum services and in the evaluation of the curricula which are so generated.[1] He distinguishes between two major modes of curriculum development as follows:

> "One mode involves system determination of guidelines with school-based determination of the detailed program of implementation. The other mode involves school-based curriculum development."

Under Mode 1 it may be necessary to set up an independent evaluation unit in order to ensure a sufficiently detached point of view.

> "Many aspects of such evaluation ought not to be in the hands of those who are responsible for their development, in order to maximise objectivity and minimise bias in the evaluation."[2]

Under Mode 2 the system's service agencies should play a supportive role.

> "This can be provided through in-service education in evaluation and through the availability of resource persons who can assist, on request, in the evaluation process itself. If these resource persons are drawn from an agency which has no administrative control over the schools and teachers, there is every chance of useful collaboration being established."[3]

There is a very important reservation about all this. Although the system authorities have moved away from their earlier practices of complete prescription in favour of prescription in more general terms there is still a powerful constraint exercised by the Boards of Secondary School Studies which conduct external examinations in most of the States.[4]

In Victoria the only external examination is that conducted for university entrance purposes at the end of Year 12.[5] Curriculum evaluation at Grade 12 level is in terms of perceived suitability as a preparation for university studies. Many other educational institutions make use of the results of the Year 12 examinations for the selection of applicants for admission to their courses. The Public Service makes use of these results for both admission to, and promotion within the various ranks of the service. A similar situation is to be found in the private sector. Despite this, there is no curriculum evaluation of the Grade 12 courses taught in most High Schools, Independent Schools and Catholic Schools in Victoria, other than that done by the Universities for their own purposes.[6]

From one point of view, the elimination of external examinations and other devices for control of the curriculum from all but Year 12 has given the schools freedom of action during Years 7 to 11 (inclusive). On the other hand most schools and most parents appear anxious to "keep the options open" by not departing too radically from the academically oriented courses that lead eventually to university entrance. The backwash of the final examination is seen to affect the curriculum of every preceding year.[7] It is for this reason that a realistic account of curriculum development and evaluation at Grade 12 level in Victoria must describe it in terms of conflict and struggle, with the various protagonists battling it out for the ascendancy of their preferred educational philosophy.

## The Grade 12 Curriculum as a Battleground

Although the Universities were the traditional custodians of the Grade 12 curriculum and any evaluation tended to be in terms of suitability as a preparation for entry to the existing university courses, this situation has been under attack in recent years. Other organisations have increased their power and influence and have become more confident and more vocal with the passage of time. Because of the insatiable appetite of the popular news media for the sensational, an atmosphere of multi-sided attack on the "élitists" has been created and maintained. There is now a wide-spread belief that the bastions of traditional education are about to crumple and fall under the attacks of progressivism. The Victorian Universities and Schools Examinations Board (V.U.S.E.B.) will cease to exist at the end of 1978. With it will go the Victorian Higher School Certificate examination (H.S.C.) and the elaborate system by which most of the places at Universities and Colleges of Advanced Education were offered to applicants on the basis of their performance at that examination. The general belief appears to be that when this happens the schools will be free, the teachers will be free and the individual students will be free.[8] There is an anticipation of something analogous to Bastille Day with its liberation of the prisoners of the "ancien régime".

The view put forward by this paper is that the tensions, rivalries and hostilities, will not fade away with the passing of the V.U.S.E.B. It is much more likely that the simplistic beliefs of those who think in

-over

terms of overthrowing a "bad" institution at all costs will be exposed and their resentments will be sharpened by their disillusionment. Curriculum development in Victoria will be closely linked with the ebb and flow of battle, with the triumphs and disasters of the contestants, and with the groupings and re-groupings which will inevitably occur as the action proceeds. Conventional methods of evaluation have obvious limitations under these conditions. Although it is tempting to believe that the role of the researcher is to provide the proper means of resolving conflicts there is no inherent reason why the perspective of such a person should be accepted. A better interpretation of current events is to regard them as the beginning of a new kind of innovation game. Surrounding the warring factions battling for the furtherance of their own vital interests is an unseen cloud of witnesses who are characterised by a strong disinclination to accept any large-scale program which is "untried". They feel that any such commitment would be irrational. They are the embodiment of that dynamic conservatism which is said to operate to resist change with a force proportional to the radicalness of that change.[9] In this context the innovation game is the resolution of the conflicts in the domain of power and prestige whilst providing an after-the-fact justification for the decisions taken by an appeal to the techniques of research and evaluation. Thus there develops a game-within-a-game whereby the myth of rational decision-making is maintained. There are two major forms which this may take. The first simply treats rational inquiry itself as a cover or smoke-screen for resolutions achieved through dominance. The second is an appeal to ideology or "true belief", which is theory held as right, inherently and once-and-for all. It makes great use of catch-cries and slogans. More importantly, it pins sinister labels on its opponents by the use of words such as, 'élitism', 'sexism', 'racism', etc. Ideologies vary enormously from the extreme left to the extreme right but they all have the great attraction of helping their followers to handle complex situations simply. In some situations these two modes may both be invoked. Ideology may demand power play to impose conformity on its non-adherents. Likewise dominance requires a theory around which to exercise power and that theory, if it is imposed rather than arrived at by rational inquiry, becomes ideology.[10] The one complements and reinforces the other creating a kind of impregnable fortress where mere intellect is at the end of its tether. It has no place except to provide an air of respectability for procedures which are not what they seem.

The danger of becoming embroiled in this kind of situation does not arise from stupidity but from the sheer incompatibility of the perspectives of the major participants. This has been described as the 'Rashomon effect' and it gives rise to much uncertainty in real-life situations where the nature of the public problem appears to different participants in different and often incompatible ways. It has been illustrated by the battleground dilemmas of Napoleon as described in Tolstoy's "War and Peace". The point of the illustration is that, "The implementer of public policy, like the commander-in-chief, works in a situation whose inherent ambiguity expresses itself through the incompatible perspectives of the key actors. ---------- each form of resolution requires some commitment to action in an uncertain situation before the formulation of a clear picture of the nature of the problem. Diagnosis comes about through intervention".[11]

The Perspectives of the Major Participants

1. The universities have a unique perspective of 120 years of complete and final authority. Although the schools have had substantial representation both on the old Schools Board and since June 1964, on the Victorian Universities and Schools Examinations Board, there has never been any doubt about the repository of power. Before any major change could occur, such as the introduction of a new subject, the approval of the universities had to be obtained at all levels, collectively, individually and by faculties and departments. Thus, in the protracted battle over Home Economics, the earlier drafts of the proposed syllabus were rejected on the grounds of insufficient academic content. When this defect was finally overcome the grounds of rejection were changed to the lack of teachers with suitable qualifications in such a variety of academic disciplines.[12] After much negotiation and some modification the subject was finally approved and suitable announcements were made.[13] However, the matter did not rest there. Certain faculties and departments at one university announced that applicants for admission to their courses would not be allowed to count the subject in their quota scores, even as an extra.[14] The criteria for this type of curriculum evaluation are far from explicit and are certainly highly subjective. Only those who have had first hand experience of the struggle can know how appropriate it is to regard the whole area of curriculum evaluation at this level as a battle-ground.

2. The colleges of advanced education have a very different perspective from the universities. Since their inception they have aspired to parity of prestige. This they saw as attainable through comparable salaries and conditions for staff and comparable entry requirements for students, to those of the universities. In the days when they were first created from the former technical colleges and teachers colleges there was much

concern that their public image should no longer be inferior to that of the universities. With the success of the major steps taken to upgrade the colleges has come a growing self confidence and a desire to pursue their own vital interests rather than look to the universities for leadership. Their adaptability and flexibility in developing new courses to meet the needs of a rapidly changing society has given encouragement to similar developments at the secondary level. Curriculum evaluation based on the process of moderation, which involves constant consultation between representatives of each college and its satellite schools, has made great progress.[15] Difficulties may arise when the moderated assessments are taken to another college for admission purposes. If such a student wishes to gain admission to a university an exemption certificate must be obtained which will exempt him from the necessity of satisfying the usual university entrance requirements by obtaining Grade D or better in English and at least three other subjects at the one H.S.C. examination.[16] This situation has given rise to much tension and dissatisfaction because, for historial reasons, the satellite schools of a college of advanced education are all technical schools. The high schools do the academic type course leading to university entrance via the higher school certificate examination. Within the technical schools there has always been a strong vocational bias and the students are divided into various streams such as professional, commercial, technician and trade courses. At the beginning of Year 12 the successful members of the professional stream may choose to enter a "tertiary orientation programme" (T.O.P.) which may be conducted at the school and moderated by consultation with the appropriate C.A.E., or may take the form of a "preliminary year" conducted at the C.A.E. A third possibility is for the T.O.P. to be conducted at a Technical College which also provides a variety of "middle level" and "sub-professional" courses for its students. Technical school students who do the T.O.P. and qualify for entry to a degree course at a C.A.E., may be recognised as having satisfied the university entrance requirements if they apply for admission to a university course leading to the corresponding degree. This is a concession which has only recently been won from the universities after much hard bargaining for reciprocal recognition. The regulation applies only to technical students since the T.O.P. is operated by the Technical Division of the Victorian Education Department. This has served to aggravate the tensions which have existed between the Technical Division and the Secondary Division of the Department for the past sixty years. Alternative Schools established under the Secondary Division sought access to tertiary education for their students by finding a C.A.E. which was willing to admit their students on the same basis as T.O.P. students, thus by-passing the H.S.C. examination.[17] This was a satisfactory means of entry to the rather restricted range of courses offered by that institution but was not much help to those who wanted admission to a university course. The Victorian Universities Admissions Committee (V.U.A.C.) ranks such applicants at the bottom of the list as E-type (unqualified). A few of these students gained entry to universities and other tertiary institutions last year by direct application coupled with a willingness to accept places which became available at the last moment due to the withdrawal of the students to whom they had been allotted by the normal V.U.A.C. procedures. The general area of conflict between the universities and the colleges is marked by numerous subsidiary battles between the C.A.E.s which originated from Technical Colleges and retain something of their former practical emphasis and those which originated from Teachers Colleges and have sought to become mini-universities. Closely related to this is the rift between the Technical and Secondary Divisions of the Education Department which gives rise to those different perspectives which make co-operation difficult. Deeply entrenched positions of independent action become sacred rights to be defended at all costs and attempts at unification, rationalisation or reconciliation only generate suspicion and hostility. These differences are also entrenched in the corresponding teachers organisations. The Victorian Secondary Teachers Association (V.S.T.A.) focuses perspectives which differ in many important respects from those of the Technical Teachers Association of Victoria (T.T.A.V.). They may at times make common cause against their major enemies, the Department, the Universities or the Colleges, but for the most part they are concerned with vital interests which are perceived very differently.

3. The teachers organisations most concerned with the Grade 12 curriculum are the Victorian Secondary Teachers Association (V.S.T.A.) and the Victorian Association of Teachers (V.A.T.). The latter arose from the "anti-ballot group" of the V.S.T.A. which became a significant opposition during the years when the official policy of the organisation was to press for the abolition of the H.S.C. examination and in those cases where the number of applicants for tertiary places exceeded the number of places available to press for the selection to be carried out by means of a ballot. The tensions generated by the controversy between the "balloters" and the "anti-balloters" spilled over into other issues and the rift continued to widen despite the removal of the balloting proposal from the offical V.S.T.A. policy.[18] The break finally came in 1976, when the V.A.T. was

-over

cont'd.

established as a separate organisation with its own distinctive policies and rules.[19] This was not just a matter of traditionalists versus progressives. The continuing struggle between the two groups is not over the need for change but over the best educational strategies to pursue in a world where change is inevitable. Both see the Grade 12 curriculum as a battleground but to the V.S.T.A. the objective is to destroy completely and finally the influence of tertiary institutions over the secondary curriculum. They think in terms of a uni-lateral declaration of independence with all secondary schools pursuing their own educational goals regardless of the expectations of the tertiary institutions. They are opposed to the selection of students for tertiary entrance by external examinations, or by teachers assessments, or by content-free tests. Their policy amounts to a demand for sufficient tertiary places to be provided so that all adults may receive the tertiary education of their choice in the institution of their choice.[20] Since 1970 they have demanded the abolition of the H.S.C. examination regardless of whether anything was available to replace it. They have seen this as the essential first step which must precede any far-reaching reform of the secondary curriculum. The V.A.T. encompasses a variety of views about the transition from secondary to tertiary education, all of which involve co-operation in order to make the passage as smooth and efficient as possible.[21] The T.T.A.V. has shown reluctance to become embroiled in the issue and appears to be reasonably satisfied with the operation of the T.O.P. although somewhat indignant at the failure to secure complete reciprocal recognition between this qualification and the H.S.C. The Victorian Teachers Union is by far the largest teachers organisation in the State but the greater part of its membership is drawn from the Primary Division. From its point of view the battle over the Grade 12 curriculum is not one that calls for any particular action in order to advance the vital interests of its members. Organisations of teachers in Independent Schools or in Catholic Schools also have primary teachers as the majority of their members. They do not become officially involved in any controversy over problems of transition from secondary to tertiary education, but those members who are concerned are usually willing to give much time and effort in the cause of co-operation, compromise and the pursuit of constructive solutions for these problems. It is worthy of note that at least some of the larger Independent and Catholic Schools have working arrangements with nearby C.A.E.[8] or Technical Colleges whereby the less academically minded of their senior students can attend vocationally oriented courses whilst still doing some subjects in their own schools. Such students have the usual access to "middle level" or to "tertiary" courses in the institutions concerned and the arrangement works very well. The teachers in the school concerned do not feel the same pressure as they would if their students were restricted to academic subjects which seemed unrelated to their real needs. This may contribute something to their lack of militancy and their willingness to co-operate with the tertiary institutions in finding constructive solutions to their students' problems. Teachers in Government schools which come under the control of the Secondary Division of the Education Department have not been nearly so successful. Eight such schools with the active support of the V.S.T.A., have developed their own alternative Grade 12 courses which they offer as equivalent to the H.S.C.[17] Since the schools do not come under the Technical Division their courses cannot be recognised as T.O.P. courses although individual C.A.E.[8] may be prepared to accept them. This move has been seen by its sponsors as a method of by-passing the H.S.C. which could undermine and rapidly destroy the whole structure of university entrance procedures. Its failure to operate as planned has led to more intense attacks on the bastions and a general increase of tension throughout the whole "broad field of battle".

4. The politicians have tended to see the situation as one of such confusion that there is not much political mileage to be gained by becoming embroiled. Certainly the V.S.T.A. has tried to enlist the support of the Opposition for its cause and some Opposition candidates at State elections have listed the abolition of the H.S.C. among their objectives, but there has been little understanding of, or enthusiasm for, the possible consequences of such an action at Government level. On the other hand, the Universities and Colleges have compelled the Minister for Education and the Cabinet to take decisive action towards the solution of the problem by setting up a new Statutory Authority with adequate powers and adequate financial resources to deal with all problems of transition from school to further education or to employment. This came about because the Victorian Universities and Colleges Committee (V.U.C.C.) set up a committee of inquiry which was called the Committee on Arrangements for Secondary Courses and Assessments (C.A.S.C.A.). The need for such an inquiry had become increasingly evident after the discovery that the V.U.S.E.B. had no legal powers except that of "conducting those examinations the passing of which is one of the conditions for entry to a university". The activities of the V.U.S.E.B. had spread far beyond this narrow interpretation of its role and it was providing a wide variety of "advisory services" which were widely considered to be an essential part of its proper functioning. Its Research Section was also carrying out experiments on the use of other methods of

assessment instead of the traditional three-hour examination. They constituted legitimate types of research, but lacked point if they could never be put into practice. They also suffered from severe constraints due to lack of finance. At that time the Board's only source of income was the examination fees paid by the candidates. There seemed no way by which changes in the Statutes of the V.U.S.E.B. could solve these problems.[22] The Universities could only delegate to the Board powers which they already had and they were too pre-occupied with their own financial problems to consider making large-scale contributions to its funds. As expected, the C.A.S.C.A. inquiry recommended the creation of a new Statutory Authority which would take over the existing functions of the V.U.S.E.B., but with the necessary powers to deal with all types of transition from school to further education or employment and vice versa.[23] The Premier, the Minister of Education and the other members of the Cabinet showed great skill in steering the necessary Bill through Parliament. The original recommendation for a widely representative Council of 61 members became a Council of 25 members with a paid Chairman, 20 nominees of the Minister of Education and 4 members of Parliament.[24] The V.S.T.A. had taken the view that the legislation was neither necessary nor desirable and when the Bill was already well on its way through Parliament it sent a deputation to the Minister to argue for a properly representative V.I.S.E. Council. This move may have had some effect, for when the Minister announced the names of his nominees it was clear that he had chosen them to represent a wide cross-section of interests. However, the V.S.T.A. was far from pleased and complained that only four practising teachers had been appointed to the V.I.S.E. Council. The V.S.T.A. President stated, "It now appears almost certain that V.I.S.E. will lead to future confrontation. There is no basis for confidence in V.I.S.E. and every reason for distrust".[25] "Meanwhile, back in the trenches," [26] the V.S.T.A. continued to develop its own campaign by organising a group of eight schools to issue their own Sixth Form Tertiary Entrance Certificate (S.T.C.). They also published a Handbook setting out details of the alternative courses offered by these schools.[27] These schools constituted the S.T.C. Group. In four of them, the S.T.C. was offered as an alternative curriculum to the H.S.C. in 1977. In the other four the H.S.C. was not offered at all. Only the S.T.C. was available. The task of actually gaining admission to courses in tertiary institutions for students who have done the S.T.C. involves a tremendous amount of personal lobbying by the teachers concerned. The R.M.I.T. alone has 65 Admissions Officers for its various courses![28] The sheer determination and willingness to undertake an enormous volume of work for very little result is fascinating to contemplate.

## A Revised Perspective?

Naturally, the universities remain unimpressed, although they are quite happy to take the odd student on an irregular entry if it happens to suit them. It is their unquestioned right to do so, but they have never shown any desire to carry out large numbers of admissions on the basis of individual negotiations. What they have done is draw up some fairly comprehensive guidelines for future admissions policy which could be taken as a blue-print for decision-making by the V.I.S.E. Council. The latter will find it very difficult to follow an independent line in developing its policy. So far, no official announcements have been made except a statement that the H.S.C. will continue in 1978 and probably in 1979 with only minor changes along the lines already planned, i.e. a greater use of teachers' assessments as part of the total H.S.C. result. Officially, this was to give the V.I.S.E. more time to consider and to develop a comprehensive policy.[29]

In the meantime, the Final Report of the Inter-University Committee on University Entrance has been published. At the time of writing the recommendations have been ratified by the Professorial Board of the University of Melbourne but have yet to be considered by the University Council.[30] The recommendations are as follows:

1. The universities would welcome the development by the V.I.S.E. subject committees, in collaboration with the schools of procedures for internal assessment of all subjects.
2. External examinations should also be retained for the present in most subjects recognised for university entrance.
3. Where there is a school component in the selection scores of subjects recognized for university entrance, the external examination could be used to scale the school marks in those subjects. V.I.S.E. should be encouraged to develop and apply procedures for establishing equivalence between marks from different schools in the State.
4. The school assessments in subjects recognized for university entrance need not be based on formal examinations or on the whole of the curriculum; it is anticipated that schools will foster initiative and experiment in this area.

-over

cont'd.

5. It is considered that the weight of the school component should be at least 20 per cent in the first instance, but consideration should be given to the desirability of moving towards a greater weighting in the light of subsequent experience.
6. Normal candidates for university entrance would present for examination in at least four subjects including English. University selection would be based on a quota score derived from the best four subject marks, plus 50 per cent of a fifth subject mark, plus 10 per cent of a sixth subject mark. This pattern of determining a quota score is intended to encourage candidates to undertake a wider range of subjects than the required minimum of four.
7. The V.U.S.E.B. subjects for the current H.S.C. Examination, with existing restrictions on subject combinations, would in the first instance, be counted in the best four subjects. It is anticipated that the list of subjects recognized as suitable for inclusion in candidates' best four subjects would be reviewed from time to time. New subjects not admitted to this list would, nevertheless, be recognised for inclusion as fifth and sixth subjects. In addition, from time to time as occasion warrants, certain subjects whose curricula are wholly designed and assessed within the schools would be included as fifth and sixth subjects.
8. Minimum qualifications for normal university entrance would, in the first instance, be those currently specified.
9. Individual faculties and schools within the universities might wish to review their prerequisite requirements for entry into their courses in the light of such new bases for selection.

The first five represent no departure from the normal practice of the V.U.S.E.B. in recent years. In six there is a change from 10 per cent to 50 per cent of the fifth subject mark. The effect of this is likely to be minimal since most students do five subjects and there is no additional incentive for them to do a sixth beyond the existing 10 per cent bonus. The last sentence in number seven does represent a new development, but it falls far short of the expectations generated by the C.A.S.C.A. Report. If this is what the Universities really want it was hardly necessary for them to destroy the V.U.S.E.B. in order to achieve it. The Report of the Inter-Universities Committee spells out very clearly what the Universities expect of the V.I.S.E. The big question is whether the new authority will continue this policy after 1979. If it does the V.S.T.A. will probably re-introduce their boycott of the H.S.C. and step-up their campaign for the widespread acceptance of the S.T.C. If it does not, it faces the possibility that the Universities will restore the V.U.S.E.B. with adequate powers to carry out the sorts of examinations, tests and assessments which they now say they want. Should this happen the V.I.S.E. may find itself devoid of influential friends, peddling a Year 12 certificate for which there is no discernible market. Such are the constraints upon rationality in the work of curriculum development and evaluation at Grade 12 level in the State of Victoria.

## Concluding Remarks

A possible model for use in a situation of conflicting perspectives has been outlined in the earlier part of this paper. The reader is invited to outline in dynamic terms the resolution of the conflicts, making use of this model or of some other model which he considers to be more suited for the purpose. There is every indication that this is an exercise which will require the utmost skill and ingenuity. It is a little different from the accepted rational experimental model of research, yet it is undeniably real and the way it is resolved will probably affect the future of education far beyond the boundaries of Victoria.

---

## REFERENCES

1. McGaw, B. Curriculum Services in a State Educational System. Education Department of Victoria, Melbourne, 1977, pp. 35.
2. Ibid., p. 16.
3. Ibid., p. 19.
4. Ibid., p. 7.
5. V.U.S.E.B., Circular to Schools, No. 53, March, 1972. "Abolition of the School Leaving Examination".
6. Rowlands, R.G., "The Outcomes of Secondary Education" in The Educational Process, The Australian College of Education, 1973, pp. 115 - 122.

7. Victorian Universities and Colleges Committee, *Report of the Committee on Arrangements for Secondary Courses and Assessment*, Section 31, p. 9, 1974.
8. Warneke, R., "Lingering Death for That Exam," *The Age*, Melbourne, 3rd September, 1974, is one example among many.
9. Rowlands, R.G., "The Chief Barrier to Large--Scale Life-Long Learning" in *Towards the Learning Society*, The Australian College of Education, 1975, pp. 75 - 82.
10. Schon, D.A., *Beyond the Stable State*, Pelican, 1973, p. 213.
11. Ibid., p. 199.
12. Minutes of the Curriculum Review Committee of the V.U.S.E.B., 1971 to 1975 inclusive.
13. V.U.S.E.B., *Circular to Schools*, No. 52, February, 1972, Supplement on Home Economics, Revisions, No. 66, June, 1974, No. 69, February, 1975, No. 70, May, 1975.
14. Ibid., No. 72, August, 1975, p. 6.
15. Stone, B., "H.S.C. is not the only Key to Tertiary Study Gateway," *The Age*, Melbourne, 15th August, 1977, p. 17.
16. V.U.S.E.B., *Circular to Schools*, No. 81, July, 1977, "Alternative Sixth Form Courses," p. 1.
17. Downes, S., "Wedge's thin end firmly in place," *The Age*, Melbourne, 16th August, 1977, p. 16.
18. Jenkin, R.E., *The Secondary Teacher*, 3, 1977, p. 7.
19. Victorian Association of Teachers, *V.A.T.*, 1, 1, August, 1976.
20. V.S.T.A., A.G.M. Report, "Open Tertiary Entrance Reaffirmed," *The Secondary Teacher*, 11, 1976, pp. 4 & 14.
21. Victorian Association of Teachers, Policy Survey, August, 1976.
22. Rowlands, R.G., *Survey of Advisory Services in Relation to the Statutory Powers of the V.U.S.E.B.: A Report to Members of the Board and its Committees*, V.U.S.E.B., Melbourne, 1971.
23. Victorian Universities and Colleges Committee, *Committee on Arrangements for Secondary Courses and Assessment: Report*, 1974, pp. 17 plus Appendices.
24. Rowlands, R.G. "Can the Community Learn?" The Australian College of Education, Eighteenth Annual Conference, May, 1977, pp. 5 - 8.
25. V.S.T.A., *The Secondary Teacher*, 3, 1977, p. 7.
26. Ibid., p. 7. This is a heading from the same article. The language of conflict is not uncommon in this journal, e.g. "______ the battle may extend well into the life of V.I.S.E. by means of boycotts or other appropriate direct action." No. 19, 1975.
27. Schools Sixth Form Tertiary Entrance Certificate Group, *Handbook, 1977*, V.S.T.A., Melbourne, pp. 217.
28. Reid, P., "H.S.C. Alternatives: The Need to Negotiate," *The Secondary Teacher*, 4, 1977, p. 5.
29. V.U.S.E.B., *Circular to Schools*, No. 81, July, 1977, p. 1.
30. University of Melbourne, *Staff News*, 5, 6, August, 1977, p. 71.

## SATURDAY 12 NOVEMBER

## 11.45 - 12.30pm

| PAPERS | Rooms |
|---|---|
| OMITTED<br>M. Carss (University of Queensland)<br>Evaluation of Mathematics Text as an Integral Part of Curriculum Planning and Evaluation | Ladies Drawing Room |
| F. Coulter (University of Western Australia)<br>G. Elsworth (Melbourne State College)<br>Teaching Behavioural and Professional Self-Perception in Student Teachers | Seminar Room A |
| N. Roseth (University of N.S.W.)<br>Empirical Evaluation of Open Education in Primary Schools | Seminar Room B |
| J. Finkelstein (Monash University)<br>Symbolic Research in the U.S.A.: Suggestions for Research | Gateway Room |
| OMITTED<br>R. Case (Ontario Institute for Studies in Education)<br>The Importance of Developmental Psychology for the Design of Effective Instruction<br>(Dr. Case will be invited to speak on this topic.) | Seminar Room D |

# TEACHING BEHAVIOUR AND PROFESSIONAL SELF PERCEPTIONS OF STUDENT TEACHERS

Frank Coulter
University of Western Australia

Gerald Elsworth
Melbourne State College

Paper presented at the Conference of the Australian Association
for Research in Education, Canberra, November 1977

## Introduction

Teacher educators are confronted by numerous difficulties in any attempt to assess the systematic effects of their programmes on student teaching behaviour. Apart from the fact that students have diverse backgrounds and entry skills, and are placed and observed in widely differing teaching contexts, institutions must rely largely upon the ratings provided by teachers and visiting 'methods' staff. These supervisors are often untrained in classroom observational skills, may use different evaluative criteria, and may hold professional values which cause them to focus upon different dimensions of classroom behaviour. Moreover, they may be required to use rating scales which call for high levels of inference about what really happened or was intended to happen in the classroom, and which permit a variety of interpretations of similarly designated behaviours. In summary, the observation and monitoring of systematic change in the teaching behaviour of students is a difficult and costly task.

Despite some of these difficulties, the present research employed direct observation in an attempt to describe the nature and extent of systematic change in the classroom behaviour of a small group of student teachers. Forty students were observed on two occasions in the year in which their first sustained teaching practice occurred, once at the beginning of the year and again toward the end of the year.

As well as attempting to describe any systematic change in the teaching behaviour of the group, the study investigated the extent to which actual classroom behaviour was related to corresponding dimensions of professional self perception. This second question was considered important because self perception is generally argued as having an interactive relationship with actual behaviour; as both affecting and being influenced by performance. If this is so, an observer's ratings of what student teachers are actually doing in the classroom should relate closely to students' perceptions of those behaviours.

A third question concerned the extent to which student teachers distinguish between their global professional self concepts and their perceptions of self in specific teaching contexts. Buchholz (1975) reported that student teachers' perceptions of certain dimensions of their teaching behaviour were unstable across different teaching situations. The existence of such instability calls into question a widely held assumption that student teachers have a baseline view of professional self which arises from the 'sum' of their more specific self perceptions. It suggests that various teaching contexts may be so radically different, and student teachers' self perceptions in these situations so varied, that the construct of *global* professional self concept may have limited explanatory power. The present investigation afforded the opportunity of investigating this question because student teachers were accessible immediately after they had taught in specific contexts. The same students also completed a global self concept measure at the beginning and end of the year.

The research reported in this paper therefore sought to investigate three separate but related questions:

(i) Is there systematic change in the teaching behaviour of students during the year in which initial practice teaching occurs?

(ii) What is the relationship between this observed behaviour and students' professional self perceptions?

(iii) Does global professional self concept relate to and reflect students' self perceptions in more specific professional contexts?

## Subjects of the Study and their Programme

The subjects were forty Bachelor of Education students, of whom twenty-seven were female and thirteen were male. They comprised ninety-three percent of the full time students in the third year of the programme concerned. Thirty-six of the students were aged from nineteen to twenty-three years, four were over twenty-three years. These students were enrolled in the third year of a four year programme designed primarily as a preparation for secondary teaching. In the first two years these students had studied Education and non-Education units concurrently, together with some limited observation in the schools. In the year in question they commenced practice teaching in secondary schools and continued their general studies in Education together with a third year non-Education unit. One of the Education units focused upon the study of teaching and classroom behaviour and involved the analysis of video tapes recorded during the two three-week practice teaching blocks. The concurrent study of the theoretical and practical aspects of teaching, preceded by two years of their Education degree was a structure which was expected to produce some systematic changes in the classroom behaviour of the group.

## Procedure

Each student was observed on two occasions by a trained classroom observer. A forty-five minute lesson was observed in the first school term; a second forty-five minute lesson was observed approximately five months later.

Ratings on eleven classroom behaviours were made at fifteen minute intervals. The classroom situations in each case were matched as closely as possible in terms of subject area, grade level, class size and the presence or absence of a supervising teacher. However, the extent to which matching of this kind was possible was recognised as one of the major limitations of the study.

Throughout each lesson the trained observer used a set of classroom observation scales (Emmer, 1970) to rate the occurrence of eleven classroom behaviours. Emmer generated his scales from existing classroom observation systems. These were OSCAR 5 (Medley and Mitzel, 1968), FAIR (Fuller, 1970), CASES (Spaulding, 1970), DYADIC (Good and Brophy, 1970) and CCS (Emmer and Albrecht, 1970). Following a series of classroom observations, intercorrelations between categories in each of these systems were factor analyzed and the resulting factors intercorrelated across systems to produce common factors (Emmer and Peck, 1974). The scales measuring Task Orientation, Clarity and Enthusiasm were not empirically based in this way, but were introduced because reviews of teacher effectiveness (Rosenshine, 1971; Rosenshine and Furst, 1971) indicated that such behaviours were related to pupil achievement. An additional scale measuring teacher reactive behaviour was generated from the research of Taylor (1972) and Sager (1974). The reactive behaviours included those which were regarded as having high learning potential for pupils.

All were five point scales and focused mainly on teacher rather than upon the pupils; they noted both verbal and non-verbal interactions and included both cognitive and affective aspects of classroom processes. The following brief account of the scales does not permit the inclusion of Emmer's examples of behaviour within the various categories which were provided for users. Severe negative affect, for example, was described as involving the display of "considerable emotion which may include physical contact, reddening of the face or a very loud voice. The event captures the attention of the whole class and disrupts the activities of the non-participants." Examples of this kind were included by Emmer to reduce the level of observer inference, particularly on the affective dimensions. Briefly then the scales were:

1. *Teacher-initiated problem solving.* This scale indicated the degree to which the teacher addressed questions and problems to the class as a whole. A rating of (1) was given to a situation where no problem solving was initiated by the teacher; a rating of (5) was given where over one-half the behaviour was teacher initiated problem solving.

2. *Pupil-to-pupil interaction.* This scale measured substantive pupil talk where pupils spoke amongst themselves about the lesson material or responded indirectly to a teacher's question. The scale ranged from (1) where there was no pupil-to-pupil interaction and where all behaviour involved routine responding, following directions, and individual seatwork, through to (5) where pupils discussed their work together for more than half the time.

3. *Teacher presentation of information.* This rated the amount of time the teacher spent presenting substantive information, either verbally or non-verbally to the class. No pupil response was required. This included straight lecturing, answering questions and writing on the blackboard. The scale ranged from (1) where teacher presentation occurred for less than twenty percent of the time through to (5) where it occurred for over eighty percent of the time.

4. *Negative affect.* This measured negative or hostile feelings shown either by the teacher or the pupils. The intensity of affect was taken in consideration. The ratings ranged from (1) where no negative affect was observed to (5) where several instances of mildly negative affect, or one or more instances of severe negative affect were observed.

5. *Positive affect.* The scale rated teacher behaviour that indicated support of, or positive regard for pupils and their behaviour. The intensity of affect was taken into consideration. A rating of (1) indicated the absence of positive affect, although not necessarily the presence of negative affect. A rating of (5) indicated three or more displays of support or regard in each fifteen minute interval.

6. *Higher level cognitive pupil behaviour.* This scale focused on pupil's verbal behaviour which evidenced thinking at high or complex cognitive levels. Such behaviour included pupils' generalizations and inferences, explanations in which they cited rules, problem solving involving principles and rules, and definitions where they cited classes of objects or events rather than single examples. Ratings ranged from (1) where no instance was observed, through to (5) where six or more certain instances occurred in the fifteen minute interval.

7. *Convergent-evaluative interaction.* This referred to interaction in which the teacher concentrated on obtaining correct answers to questions with little or no attempt to continue the interaction after the answer was received. A rating of (1) was given where little or no interaction of this kind occurred; where more than 75% of the interaction was convergent-evaluative a rating of (5) was given.

8. *Task orientation.* This scale judged the extent to which the teacher's behaviour was related to the identifiable instructional objectives of the lesson. Ratings ranged from (1) where there was much wasted time and wandering unfocused discussion with little substantive progress attempted, through to (5) where class activities were carefully planned and the emphasis was clearly upon pupil attainment of lesson objectives.

9. *Clarity.* This scale rated the extent to which the pupils appeared to understand the material the teacher presented to them, and the extent to which the interaction between the pupils and the teacher appeared to further clarify the material. A rating of (1) was given where pupils seemed very confused by the presentation and where questions were answered in an unclear manner. A rating of (5) was given where questions were answered adequately using concepts and terms familiar to the pupils.

10. *Enthusiasm.* This scale rated the teacher's interest, vitality and involvement in the lesson. Ratings ranged from (1) where the teacher behaviour was lethargic, dull, routine, with a minimum of variability through to (5) where the teacher appeared stimulating, energetic, alert and interested in what she was teaching.

11. *Teacher reactive behaviour.* This scale measured the number of times the teacher reacted to pupils' responses by further probing the response, by extending the response with further information, or by providing the pupil with corrective information, as distinct from being non-committal or making pupil-related rather than task-related comments. Ratings ranged from (1) to (5) according to whether the teacher's reactive behaviour had high learning potential for pupils.

The scales were used by an observer who had previously used Emmer's Classroom Observation Scales in a related study (Archer, 1974). After training with video tapes and after practice in actual classroom settings, a high level of agreement had been reached with another trained observer. It was accepted therefore that the observer rated consistently and reliably. To ensure reliable use of the new Teacher Reactive Behaviour scale the observer rated ten video tapes of classroom teaching on two occasions one week apart. The stability coefficient was .82.

After each lesson the student teachers completed a scale designed to measure seven dimensions of self perception which were Clarity, Orderliness, Warmth-supportiveness, Energy, Creativity, Non-conformity and global Satisfaction (Elsworth and Coulter, 1977). The purpose was to compare these contextually specific professional self perceptions with actual classroom behaviour.

## Results

### (i) Observed classroom behaviour

Table 1 shows the mean scores on the eleven dimensions of classroom behaviour, and the correlations between the initial observations and those made five months later. The similarity between the pretest and post-test means indicates that with the exception of Clarity and Higher level cognitive pupil responses there was very little systematic change in the behaviour observed in the classrooms of the group as a whole.

The low test-retest correlations might have been interpreted as indicating that there was a considerable degree of change by individuals within the group as they searched for their own personal teaching styles. However, inspection of the raw data revealed that the generally low test-retest correlations were not a result of significant movements by individuals being offset by movement in the opposite direction by other individuals. They were rather a result of unsystematic movement by most students within a *narrow range* of scores. For example, in the case of pupil-to-pupil interaction (r=.24), thirty of the forty classrooms observed remained within the 1 to 2 range on both occasions. In fact, one-half of the classrooms observed were rated as having *no* pupil to pupil interaction on either occasion and as involving only routine responding or individual seatwork. Inspection of the data relating to teacher reactive behaviour (r=.11) revealed a very similar pattern with random movements by thirty-one of the forty students between categories 1 and 2. From this it may be concluded that there was not only a general lack of systematic change within the group, but also that the change observed in *individuals* was generally slight.

It should be observed that the preponderance of ratings in the lower categories for behaviours such as Pupil-to-pupil interaction, Teacher reactive behaviour, Enthusiasm and Higher level cognitive response did not arise from the observer's failure to utilize the full scale ranges. In each instance observations were recorded in each of the scale categories. The observer was apparently able to recognise and record extremes where they occurred.

TABLE 1

MEAN SCORES AND ASSOCIATED CORRELATED 't' TESTS ON ELEVEN DIMENSIONS OF CLASSROOM BEHAVIOUR FOR THE GROUP OF STUDENT TEACHERS OBSERVED ON TWO OCCASIONS (N=40)

| Dimensions of Classroom Behaviour | $\bar{x}$ Obs 1 | $\bar{x}$ Obs 2 | r | t | p (2-tailed test) |
|---|---|---|---|---|---|
| Teacher initiated problem solving | 3.34 | 3.05 | 0.21 | 1.02 | n.s. |
| Pupil-to-pupil interaction | 1.44 | 1.43 | 0.24 | 0.07 | n.s. |
| Teacher presentation of information | 2.40 | 2.81 | 0.37 | 1.58 | n.s. |
| Negative affect | 2.33 | 2.36 | 0.57 | 0.19 | n.s. |
| Positive affect | 3.11 | 3.19 | 0.60 | 0.54 | n.s. |
| Higher level cognitive pupil response | 1.53 | 1.31 | 0.57 | 2.10 | 0.04 |
| Convergent-evaluative interaction | 3.34 | 3.41 | 0.38 | 0.25 | n.s. |
| Task orientation | 4.24 | 4.39 | 0.25 | 1.03 | n.s. |
| Clarity | 3.24 | 3.70 | 0.55 | 3.60 | <0.001 |
| Enthusiasm | 2.31 | 2.51 | 0.48 | 1.71 | n.s. |
| Teacher reactive behaviour | 1.59 | 1.33 | 0.11 | 1.52 | n.s. |

Interpretation of these behaviour ratings must be qualified by the fact that Emmer provided no norms against which these data might be compared. What is a 'high' or 'low' incidence of a particular behaviour is therefore somewhat speculative. With this limitation in mind, the picture which emerged for the group as a whole on both occasions closely resembled the stereotype of teacher-dominated, task-oriented teaching in which considerable time was devoted to the presentation of information, where many questions tended to be convergent in the sense of confining pupils to a predetermined structure, and where pupil-to-pupil interaction was minimal.

If this is a model toward which beginning secondary teachers are being socialised then it is unsurprising that students high on social extroversion and rigidity feel happier and better adjusted within it (Coulter and Elsworth, 1976). This model is somewhat at odds with those expounded in some training institutions and schools where the emphasis is upon more open teaching and learning structures which stress pupil initiative and participation. However, the model which emerged here is familiar and pervasive, and is probably within the experience of most readers either as pupils or teachers. Perhaps it persists because it 'works' and is the one most easily adopted by beginners trying to manage twenty or thirty unfamiliar children; and it is one which, with slight variations, is practised and promoted by many more experienced teachers because they, too, see that it has value in this pragmatic sense. If this is true, then a great deal of preservice theorising about teaching and learning may be having little impact in terms of fostering new or different patterns of teaching in conventionally designed classrooms. Practice teaching experience may inform the student that much of what is said in lectures does not relate to the way teaching *is*; and perhaps it fails to provide them with strategies to make it significantly different. This may explain why a recent survey of ten different teacher education programmes (Coulter and Elsworth, 1977) revealed that between one-third and one-half of students perceived their on-campus courses as having little or no value to them as a preparation for teaching.

## (ii) The relationship between observed teaching behaviours and corresponding dimensions of self perception

The classroom behaviour data reported in the preceding section provided an opportunity for exploring the relationship between classroom behaviour in a specific context with students' self perceptions in that same context. Although some of the aspects of professional self perception did not correspond exactly to the dimensions of classroom behaviour observed, some relationships were expected. For example, one would expect a relationship between self perceptions of Orderliness and Clarity, and the observer's estimates of the amount of task-oriented teaching behaviour. Similarly, a relationship was predicted between students' perceptions of Warmth-supportiveness and the observed levels of Positive and Negative affect in the classroom. Also, students' perceptions of their Non-conformity were expected to relate to levels of observed Pupil-to-pupil interaction, Convergent evaluative interaction, Teacher presentation and Task orientation because teachers who see themselves as non-conforming would presumably be eager to discard the traditional teacher-centred, tightly structured lesson in favour of a more pupil-centred approach. It should be added that in this particular context the terms 'traditional' and 'non-conforming' are used in a value-free sense.

It will be apparent that this aspect of the study was somewhat exploratory. It sought to clarify whether there was any relationship between relatively inexperienced student teachers' perceptions of their behaviour in a specific context, and how they were seen to be by a trained observer.

Table 2 reports the correlations between the 'comparable' dimensions of behaviour and self perception. Correlations between all dimensions of self perception and all dimensions of observed teaching behaviour are presented in Appendix 1.

TABLE 2

RELATIONSHIPS BETWEEN TEACHING BEHAVIOURS AND
COMPARABLE DIMENSIONS OF SELF PERCEPTION (N=40)
(SECOND OBSERVATION IN PARENTHESIS)

| "Myself as a Teacher Today with Grade __" | Observed Teaching Behaviours | r | r |
|---|---|---|---|
| Orderliness | Task orientation | .32* | (.50**) |
| | Clarity | .31* | (.43**) |
| Warmth | Positive affect | .53** | (.41**) |
| | Negative affect | -.36** | (-.36**) |
| Clarity | Clarity | .25 | (.14) |
| Enthusiasm | Enthusiasm | .35* | (.40**) |
| Non-conformity | Pupil-to-pupil interaction | .34* | (.42**) |
| | Teacher-initiated problem solving | -.12 | (-.20) |
| | Teacher presentation of information | -.14 | (-.20) |
| | Convergent evaluative interaction | -.36** | (-.16) |

* $p < .05$ for one-tailed test
** $p < .01$ for one-tailed test

In the majority of cases the predicted relationships were significant; and in each instance they were in the expected direction. The low correlation between perceived and observed Clarity was surprising, but may have arisen from the fact that Clarity as defined by Emmer embraced aspects of Orderliness in addition to aspects of Clarity or lucidity. This interpretation is supported by the relatively high correlation between perceived Orderliness and observed Clarity (Table 2).

These results provide some support for the view that student teachers are fairly sensitive to what is happening in their classrooms. This was especially the case for Positive and Negative affect where the directions of the relationships of the observer's ratings with the students' self perceptions of Warmth were strong and clear. The results also provide further evidence of the validity of the self perception scales (Elsworth and Coulter, 1977).

(iii) The relationship between self perception in specific teaching contexts and global professional self concept

Students' self ratings of "Myself as a Teacher Today with Grade __" were completed immediately after each lesson had been observed. These perceptions of self were compared with the ratings of the more global concept "Myself as a Teacher" which were made in March and seven months later after the completion of teaching practice. The scales used to rate each concept were identical (Appendix 2).

Figures 1 to 7 show that *at the pretest* there were discrepancies between the means for specific self perception and more generalized self concept. Table 3 indicates that for each dimension except Satisfaction these differences were significant in the direction of more positive global self perceptions. However, these differences diminished on every dimension by the second observation. The only dimensions on which significant discrepancies persisted were Warmth and Clarity (Table 4).

## TABLE 3

## PRETEST: MEAN SCORES AND ASSOCIATED CORRELATED 't' TESTS FOR "MYSELF AS A TEACHER" AND "MYSELF AS A TEACHER TODAY" (N=40)

| Dimension of Self Perception | Mean: "Myself as a Teacher" | Mean: "Myself as a Teacher Today" | r | t | p (2-tailed test) |
|---|---|---|---|---|---|
| Creativity | 14.40 | 12.60 | 0.69 | 3.96 | <0.001 |
| Orderliness | 30.80 | 27.30 | 0.29 | 3.03 | 0.004 |
| Warmth | 33.15 | 28.10 | 0.64 | 5.52 | <0.001 |
| Satisfaction | 12.60 | 11.70 | 0.48 | 1.28 | n.s. |
| Clarity | 19.58 | 17.03 | 0.20 | 3.76 | 0.001 |
| Energy | 26.13 | 24.30 | 0.49 | 2.57 | 0.014 |
| Non-Conformity | 19.28 | 17.33 | 0.52 | 2.48 | 0.018 |

## TABLE 4

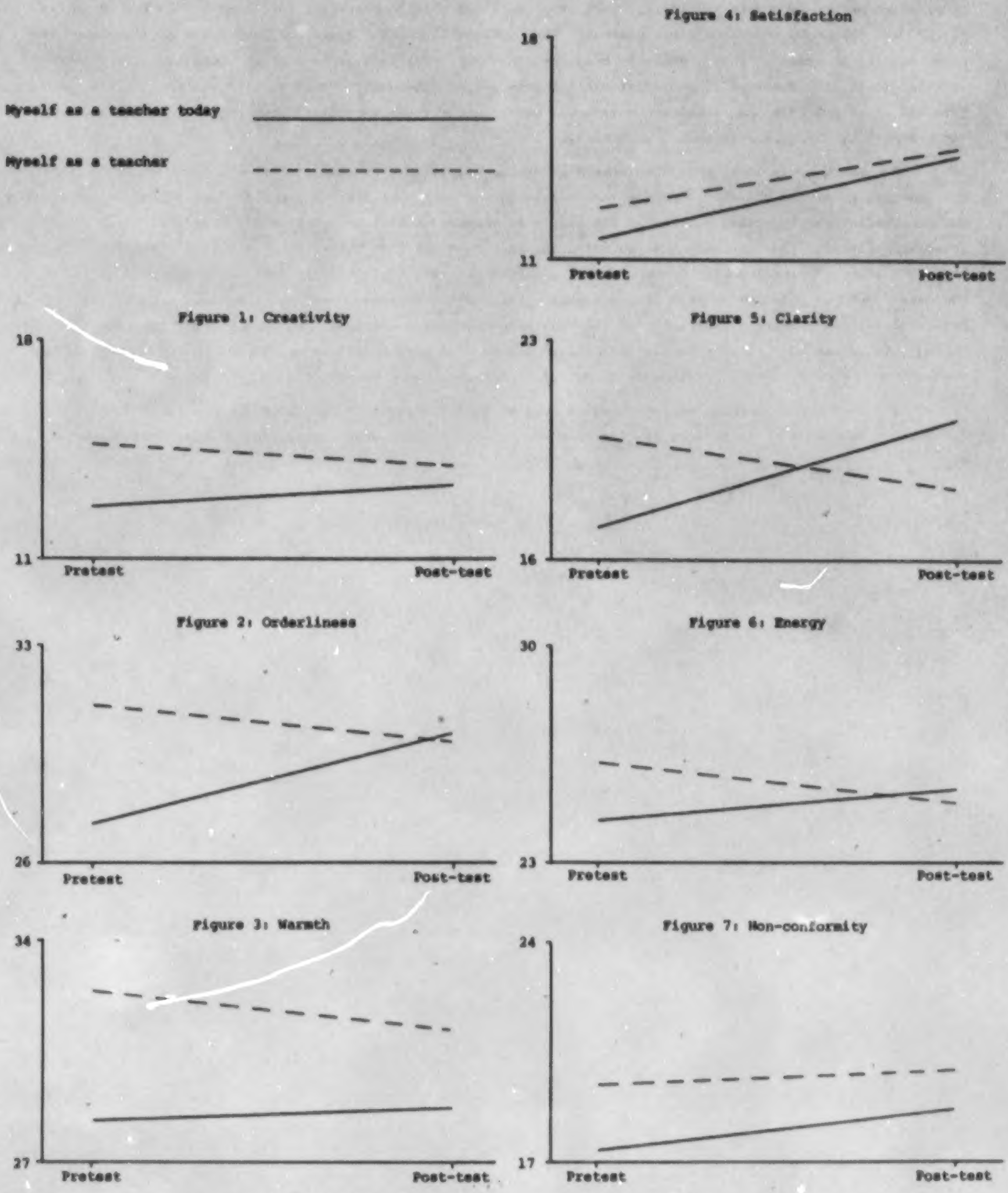
## POST-TEST: MEAN SCORES AND ASSOCIATED CORRELATED 't' TESTS FOR "MYSELF AS A TEACHER" AND "MYSELF AS A TEACHER TODAY" (N=38)*

| Dimension of Self Perception | Mean: "Myself as a Teacher" | Mean: "Myself as a Teacher Today" | r | t | p (2-tailed test) |
|---|---|---|---|---|---|
| Creativity | 13.13 | 13.82 | 0.47 | 1.36 | n.s. |
| Orderliness | 29.97 | 29.89 | 0.66 | 0.10 | n.s. |
| Warmth | 28.74 | 31.00 | 0.69 | 2.75 | 0.009 |
| Satisfaction | 13.95 | 14.21 | 0.55 | 0.45 | n.s. |
| Clarity | 20.26 | 18.26 | 0.38 | 3.20 | 0.003 |
| Energy | 25.16 | 24.95 | 0.49 | 0.27 | n.s. |
| Non-Conformity | 18.45 | 19.76 | 0.60 | 1.92 | n.s. |

* two responses were unusable

The original discrepancy might have arisen from the fact that the global self concept measure was taken prior to any teaching experience in schools. The more specific measure was recorded immediately after initial classroom teaching. Global professional self concept at the pretest may therefore have been inaccurately optimistic of future teaching behaviour. Support for this interpretation derives from the fact that for *each* dimension global self concept was higher, or more positive, than was "Myself as a Teacher Today with Grade __". The fact that this discrepancy diminished on every dimension *after experience*, suggests that professional self perception at a particular time approximates closely to global professional self concept.

FIGURES 1-7 PRETEST AND POST-TEST MEANS FOR RATINGS OF "MYSELF AS A TEACHER" AND "MYSELF AS A TEACHER TODAY WITH GRADE __" (N=40)



## Conclusion

The investigation reported here was largely exploratory. Apart from seeking to describe any systematic change in student teaching behaviour, it explored the nature of the relationship between observed behaviour and corresponding dimensions of professional self concept. The relationship between self perception in specific teaching contexts and more global professional self concept was also described.

Although no systematic change was reported for the teaching behaviour of the group, it should be recalled that this was the year in which *initial* teaching practice occurred and a total of only six weeks sustained classroom contact was involved. This may have been too short a time to effect behavioural change. It is interesting that in many other programmes six weeks represents the total practice component. Questions may therefore be asked whether much systematic change occurs in *school based* teaching prior to graduation in some programmes. This observation may be even more significant when it is recalled that both systematic change within the group *and* significant change by individuals was minimal. The findings also raise questions concerning the impact of the more theoretical strands of programmes on teaching behaviour. The profile of teaching behaviour did resemble a somewhat familiar stereotype of a traditional teacher-centred classroom. However, this was the initial practice year and it is possible that these students would experiment with different and more pupil-centred approaches at later stages of their professional development.

A comment needs to be made about the size of the sample in this study. Although forty students presented the researchers with a sizeable observational task, they offered only limited scope for generalisation. Not only was the student sample small, but so was the sample of observed behaviour. To these limitations which relate to the generalisability of the findings, must be added the fact that the live observation of eleven dimensions of classroom behaviour involved considerable abstraction. Despite the relatively high reliabilities achieved by the observer, the data should be interpreted with this methodological problem in mind. However, the results do suggest that the impact of initial classroom experience on teaching behaviour may be minimal. Perhaps changes in student teachers during this period of initial exposure to the role may be of a more covert nature involving such things as the reassessment of professional commitment, change in self image, growth in confidence, and so on.

The findings concerning self perception suggest that how students perceive themselves in specific teaching contexts is related both to the actual classroom events and to their more global professional self concepts.

# REFERENCES

Archer, J. Supervisors' perceptions of their student teachers' classroom behaviours. Unpublished honours thesis, James Cook University of North Queensland, 1974.

Buchholz, M. The stability of some dimensions of the professional self image of student teachers. Unpublished honours thesis, James Cook University of North Queensland, 1975.

Coulter, F. and Elsworth, G.R. Professional self perception and personality characteristics of student teachers. *Education Research and Perspectives*, 1976, 3, 49-61.

Coulter, F. and Elsworth, G.R. An evaluation of the impact of initial student teaching on professional self perception, commitment and teaching behaviour. Report submitted to the Education Research and Development Committee, Canberra, October, 1977.

Elsworth, G.R. and Coulter, F. *Aspiration and attainment: The measurement of professional self perception in student teachers.* Occasional Paper No. 11. Australian Council for Educational Research, 1977.

Emmer, E.T. Classroom observation scales: A manual. Research and Development Centre for Teacher Education, University of Texas at Austin, 1970.

Emmer, E.T. and Albrecht, I.D. The cognitive components system manual. The Research and Development Centre for Teacher Education, University of Texas at Austin, 1970.

Emmer, E.T. and Peck, R.F. Dimensions of classroom behaviour. *Journal of Educational Psychology*, 1973, 64, 223-240.

Fuller, F.C. FAIR system manual. The Research and Development Centre for Teacher Education, University of Texas at Austin, 1970.

Good, T.L. and Brophy, J.E. Teacher-child dyadic interaction: A manual for coding classroom behaviour. The Research and Development Centre for Teacher Education, University of Texas at Austin, 1970.

Medley, D.M. and Mitzel, H.E. Techniques for measuring classroom behaviour. *Journal of Educational Psychology*, 1958, 49, 86-92.

Rosenshine, B. *Teaching behaviours and student achievement.* Windsor: N.F.E.R., 1971.

Rosenshine, B. and Furst, N. Research on teacher performance criteria. *In* Smith, B.O. (ed.) *Research in teacher education.* Englewood Cliffs, New Jersey: Prentice Hall, 1971, 37-72.

Sager, B. Some aspects of curricula context in relation to classroom teaching processes. Unpublished honours thesis, James Cook University of North Queensland, 1974.

Spaulding, R.L. *Classroom behaviour analysis and treatment.* San Jose State College, 1970.

Taylor, S.C. A classroom study of teacher reactive behaviour. Unpublished honours thesis, James Cook University of North Queensland, 1972.

APPENDIX 1

## RELATIONSHIPS BETWEEN SELF PERCEPTION (STATE) AND OBSERVED TEACHING BEHAVIOUR (N=40)

(SECOND OBSERVATION IN PARENTHESIS)

"MYSELF AS A TEACHER TODAY WITH GRADE __"

| Behavioural Observation Rating Scales | Creativity | | Orderliness | | Warmth | | Satisfaction | | Clarity | | Energy | | Non-conformity | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Teacher-initiated problem-solving | .01 | ( .19) | .16 | ( .09) | .27* | ( .00) | .36** | ( .00) | .20 | (-.03) | .11 | ( .10) | -.12 | (-.20) |
| Pupil-to-pupil interaction | .03 | ( .27*) | .19 | (-.03) | -.03 | ( .33*) | .15 | ( .27*) | .15 | ( .12) | .01 | ( .36**) | .34* | ( .42**) |
| Teacher presentation of information | -.01 | (-.16) | .10 | (-.05) | -.11 | (-.19) | -.08 | (-.13) | -.18 | (-.02) | .12 | (-.13) | -.14 | (-.22) |
| Negative affect | .10 | (-.18) | .10 | (-.05) | -.36** | (-.36**) | -.17 | (-.26*) | -.02 | ( .13) | -.07 | (-.26*) | .05 | (-.14) |
| Positive affect | .32* | ( .41**) | .10 | ( .23) | .53** | ( .41**) | .31* | ( .24) | .04 | ( .07) | .38** | ( .31*) | .41** | ( .19) |
| Higher cognitive level pupil response | .18 | (-.07) | -.18 | (-.08) | .01 | (-.12) | .09 | (-.03) | -.02 | (-.22) | .12 | ( .02) | .24 | (-.16) |
| Convergent-evaluative interaction | -.15 | ( .02) | .29* | ( .31*) | -.10 | (-.09) | .01 | ( .09) | -.16 | ( .12) | .01 | ( .19) | -.36** | (-.16) |
| Task orientation | .20 | ( .41**) | .32* | ( .50**) | .35* | ( .35*) | .46** | ( .37**) | .21 | ( .15) | .32* | ( .59**) | .13 | (-.15) |
| Clarity | .32* | ( .32*) | .31* | ( .43**) | .30* | ( .33*) | .54** | ( .41**) | .25 | ( .14) | .51** | ( .55**) | .33** | (-.07) |
| Enthusiasm | .25 | ( .34*) | .27* | ( .33*) | .24 | ( .25) | .40** | ( .29*) | .18 | ( .16) | .35* | (.40**) | .39** | ( .12) |
| Teacher reactive behaviour - high learning potential | -.06 | ( .06) | -.36** | (-.09) | .16 | (-.02) | -.05 | (-.22) | .03 | (-.27*) | -.03 | (-.20) | -.23 | (-.05) |

* p <.05 for one-tailed test
** p <.01 for one-tailed test

APPENDIX 2

## SCALES USED TO RATE THE CONCEPTS "MYSELF AS A TEACHER"
## AND "MYSELF AS A TEACHER TODAY WITH GRADE __"

warm ______ : ______ : ______ : ______ : ______ : ______ : ______ cool

chaotic ______ : ______ : ______ : ______ : ______ : ______ : ______ orderly

comforting ______ : ______ : ______ : ______ : ______ : ______ : ______ reproaching

friendly ______ : ______ : ______ : ______ : ______ : ______ : ______ hostile

free ______ : ______ : ______ : ______ : ______ : ______ : ______ constrained

spirited ______ : ______ : ______ : ______ : ______ : ______ : ______ apathetic

lucid ______ : ______ : ______ : ______ : ______ : ______ : ______ obscure

systematic ______ : ______ : ______ : ______ : ______ : ______ : ______ random

insulting ______ : ______ : ______ : ______ : ______ : ______ : ______ esteeming

creative ______ : ______ : ______ : ______ : ______ : ______ : ______ uncreative

disorganized ______ : ______ : ______ : ______ : ______ : ______ : ______ organized

conventional ______ : ______ : ______ : ______ : ______ : ______ : ______ unconventional

indifferent ______ : ______ : ______ : ______ : ______ : ______ : ______ eager

satisfied ______ : ______ : ______ : ______ : ______ : ______ : ______ dissatisfied

vague ______ : ______ : ______ : ______ : ______ : ______ : ______ clear

prepared ______ : ______ : ______ : ______ : ______ : ______ : ______ unprepared

jumbled ______ : ______ : ______ : ______ : ______ : ______ : ______ arranged

puzzled ______ : ______ : ______ : ______ : ______ : ______ : ______ informing

restrained ______ : ______ : ______ : ______ : ______ : ______ : ______ liberated

blurry ______ : ______ : ______ : ______ : ______ : ______ : ______ sharp

fresh ______ : ______ : ______ : ______ : ______ : ______ : ______ stale

mean ______ : ______ : ______ : ______ : ______ : ______ : ______ kind

adaptable ______ : ______ : ______ : ______ : ______ : ______ : ______ rigid

enthusiastic ______ : ______ : ______ : ______ : ______ : ______ : ______ unenthusiastic

inert ______ : ______ : ______ : ______ : ______ : ______ : ______ energetic

usual ______ : ______ : ______ : ______ : ______ : ______ : ______ unusual

imaginative ______ : ______ : ______ : ______ : ______ : ______ : ______ unimaginative

efficient ______ : ______ : ______ : ______ : ______ : ______ : ______ inefficient

discontented ______ : ______ : ______ : ______ : ______ : ______ : ______ contented

conforming ______ : ______ : ______ : ______ : ______ : ______ : ______ non-conforming

rewarding ______ : ______ : ______ : ______ : ______ : ______ : ______ punishing

fulfilled ______ : ______ : ______ : ______ : ______ : ______ : ______ frustrated

Naomi Roseth, School of Psychology, University of N.S.W.

# EMPIRICAL EVALUATION OF OPEN EDUCATION IN PRIMARY SCHOOLS

## Introduction

The studies reported here were conducted within a framework of a joint programme by the New South Wales Department of Education and the School of Psychology, the University of New South Wales. The studies were designed to evaluate the processes and outcomes operating in primary school open classes. Early in 1972 the Department of Education constructed a number of open space, two-teacher classrooms in primary schools in the Sydney area. The Department's intention was to provide a physical environment which would facilitate the 'opening up' of educational methods. The building of the open space classrooms was seen at first as an experimental programme, to be evaluated independently by the School of Psychology. Four separate studies were conducted in the two schools which were the first to receive the open space units. Two were comparative in nature: educational processes and outcomes for children in the open space classrooms (experimental groups) were compared to those of similar children in conventional space classrooms (control groups). These studies were conducted in both schools (referred to as A and B). The third and fourth studies were conducted in School A. One took the form of extensive interviews with children in the experimental and control classes aimed at assessing their perception of and opinions about their classrooms; the other involved systematic observations of children in the open space classroom aimed at gaining an insight into the use they made of its special features.

## Review of Open Education Research

Open education literature tends to be impressionistic - the work of committed and enthusiastic educators and journalists (e.g., Featherstone 1971, Kohl 1969, Nyquist and Hawes 1972, Silberman 1973). Nevertheless, empirical research has also been conducted in recent years. The bulk of this research is comparative: open classes, serving as experimental groups, are compared to traditional classes, serving as control groups. However, most open education research has not come to grips with three important aspects: The operational definition of educational openness; the design of studies and the selection of measures.

The operational definitions of openness and traditionality are the independent variables of comparative studies. Researchers in the past have defined openness in terms of one of three criteria: the physical features of the classroom (e.g., Clem et al. 1974, Edwards 1973, Grapko 1972, Lueders-Salmon 1971); or scores on one of the existing openness rating scales (e.g., Corlis 1973, Franks et al. 1974, Kohler 1973, McPartland and Epstein 1975); or organizational strategies adopted by teachers (e.g., Innes 1974, Renthrow et al. 1973, Reynolds 1974, Tuckman et al. 1974). By using any one of these criteria singly, researchers have grossly oversimplified the issue of operationally defining openness and traditionality. There is no obvious reason why physical features alone should be associated with educational outcomes; teachers or observers cannot be expected to rate classroom openness objectively; the adoption of open organizational strategies does not guarantee the implementation of open teaching methods.

With regard to the design of studies, random allocation and matching of subjects in experimental and control groups are hard to achieve in educational field research. This is because the kind of schools or classes children attend is usually influenced by parents' choice. Of 37 studies conducted between 1971 and 1975, only one (Owen 1974), reports random allocation of subjects to open and traditional classes, and only eight report matched grouping with data on matched characteristics supplied (Bleier et al. 1972, Clem et al. 1971, Klein and Eshel 1973, Kohler 1973, McPartland 1973, Ruedi and West 1972, Scheirer 1972, Sullivan 1971). In the absence of random allocation or proper matching, attribution of differences between experimental and control groups to differences in educational methods is risky.

The constantly changing nature of classroom environment is generally not taken into account in research design. Most research is in the form of one-shot studies, in which measures are taken at one point in time. In ten studies measures were performed more than once (Burnham 1973, Crandall 1973, Grapko 1972, Heimgartner 1972, Killough 1971, McPartland 1972, Moddie 1971, Reynolds 1974, Sewell et al. 1975, Tuckman et al. 1974). However, the degree of openness of the experimental groups and traditionality of the control groups was determined in these studies at the outset of the research programme. The educational environment against which later measures were taken was therefore not ascertained.

As for the choice of measures, evaluative research on open education must be based on variables which reflect the philosophy and objectives of open as well as traditional education. In traditional education there is more emphasis on cognitive skills than in open education which is concerned with a broader range of pupils' attributes and attitudes (Angus 1973, Hogben 1974). Of the studies mentioned earlier, only eight measured more than one type of variables (Grapko 1972, Owen 1974, Reynolds 1974, Sackett 1971, Scheirer 1972, Sewell 1975, Traub et al. 1973, Tuckman et al. 1974). The majority measured only cognitive or only non-cognitive variables and thus failed to provide a comprehensive evaluation.

The predominant use of standardized tests for the measurement of the dependent variable is the most striking shortcoming of comparative studies. An effort was made by Bleier (1972), Corlis (1973), Renthrow et al. (1973) and Traub et al. (1973) to measure student outcome variables by the use of specially devised behavioural probe measures. All other researchers used standardized paper-and-pencil tests without allowing for the fact that these tend to favour traditional class children who are more accustomed to test-taking situations.

## The University of N.S.W. Research Programme

### Background

The two schools which were the first to receive the open space classroom within the N.S.W. Education Department experimental programme were convenient sites for research purposes. In both schools comparable classes remained in existing, conventional space classrooms and could serve as control groups. The principals and teachers involved were willing to accommodate the research programme within their schools. The background against which the innovation was introduced differed somewhat in the two schools: School A was a demonstration school in an inner suburb of Sydney. The teachers who taught in the open space classroom, although trained and experienced in traditional methods, were committed to open education and volunteered to teach in that classroom. The teacher who remained in the conventional space classroom was equally trained and experienced but committed to the methods of traditional education. School B was in an outer suburb of Sydney. The teachers in the open space classroom had no special interest in open education and were appointed by the school administration to teach in it.

The open space classrooms built by the Department of Education were composed of a large carpeted learning area, equipped with light, moveable desks and chairs, divider units and ample storage space; and a 'practical area', equipped with sinks and working benches, designed for 'messy' activities. Two classes and two teachers joined together in these classrooms as a team.

Of the two comparative studies conducted, one, in School A involved a multiplicity of measures and spanned over two years when children were in third and fourth grades. The other, in School B, was a smaller scale study and spanned over two school terms, when children were in first grade.

### Comparative Studies, Schools A and B

Definition of openness:

Openness and traditionality were defined in the context of this research in terms of three factors: the physical layout of the classrooms, the organizational strategies adopted by teachers and teaching methods. Physical layout and organizational strategies could be easily ascertained. Teaching methods were systematically observed and measured by time sampled observation techniques, devised for this research. The teaching methods measured were those which on the basis of consultations with educators and open education literature (e.g., Barth 1969, Bussis and Chittenden 1970, Stephens 1974) most commonly differentiated between open and traditional practices. The measures devised were referred to as "Teacher Engaging Class as a Whole" (T.E.C.A.W.) which measured the proportion of time within a specified period during which the teacher engaged the attention of the class as a whole, and "Simultaneous Activities Count" (S.A.C.) which measured the number of curricular and non-curricular activities occurring simultaneously at a given time (Roseth 1976). Teaching methods were observed and measured repeatedly in both schools throughout the study period.

An open class was thus defined as having open space layout; open organizational strategies (specifically: contract system, absence of timetables and of fixed seating) and open teaching methods (individual or group teaching and diversified activities). A traditional class was defined as having a conventional space layout; traditional organizational strategies (specifically: lesson system, timetables and fixed, teacher determined seating) and traditional teaching methods (whole-class teaching and whole-class activity). These operational definitions were based on the recognition that open and traditional classes are multi-dimensional, rather than definite concepts and that educational openness is a relative concept. The openness of the experimental classes and the traditionality of the control classes were not taken for granted but demonstrated empirically and not only at the outset of the research programme but on a number of occasions throughout the research.

Study design:

The different circumstances in the two schools led to different study designs. At School A three groups participated in the study: two experimental groups (E1 and E2) who moved with their teachers from conventional to the open space classroom (N = 22 and 15, only children who were present at all testing occasions over the two years were included in the analysis of results) and a control group (C) which remained in its conventional space classroom (N = 13). (The number of children present in classes E1, E2 and C at the beginning of the study was 32, 31 and 29 respectively. The marked decrease was a function of absence on testing occasions and the location of the school in an area with a relatively high population mobility.) Children were randomly allocated to classes by the school administration. Behavioural observations and tests of pupil outcomes were carried out during four testing occasions: one before the move of the experimental groups to the open space classroom and three at subsequent occasions, two, six and 18 months after the move. The random allocation of children to classes and the fact that the experimental group could be "captured" for the purpose of measurement prior to their move to the open space classroom enhanced the power of the design since it allowed for a comparison of similar groups of children before and on a number of occasions after exposure to different educational treatments.

Four groups participated in the study at School B: two experimental groups (E1 and E2) which comprised the children in the open space classroom (N = 28 and 24); a control group (C1) which comprised children who moved to a new conventional space classroom adjacent to the open space classroom in the newly renovated building (N = 27) and another control group (C2) which remained in its existing conventional space classroom (N = 20). (No significant losses or gains occurred in the number of children within one school term.) Children were allocated to classes by age so that those in each class were born within about three months of each other. Unlike the open space classroom at School A, the one at School B was fitted with a concertina door across the learning area. This enabled a conversion of the open space classroom into two self-contained, conventional space classrooms. Behavioural observations and testing of children were carried out in three stages: a baseline stage when the door across the open space classroom was closed and the two experimental groups were in two conventional space classrooms; an experimental stage when the door was open and the two groups operated in an open space classroom and a reinstitution stage when the doors were closed again. This design can be viewed both as a within-subject and as an experimental-control group design. This dual feature allowed a comparison of the experimental group to the control group as well as to itself under different conditions.

Measures:

A battery of measures comprising both cognitive and non-cognitive variables was administered in School A. Cognitive variables were measured by standardized paper-and-pencil tests commonly used by the school. The variables measured were mathematics (three subtests, M1, M3 and M5 of the N.S.W. Primary Evaluation Programme, Australian Council for Educational Research, and Department of Education of N.S.W. 1972) and reading (Silent Reading Tests, R3 and R4, Schonell 1959). The tests were administered to groups of children away from the classrooms using a standardized, controlled procedure.

The applied nature of the research programme governed the selection of non-cognitive variables. Discussions were held with teachers and administrators to establish the outcomes which they perceived as likely to be fostered by open education. Five variables emerged: inquiry, cooperation, yielding to authority, creativity and performance in teachers' absence. The first four variables were measured in School A, the last in School B.

Of these variables, one - creativity - was tested by a standardized paper-and-pencil test (the verbal form of the Torrance Test of Creative Thinking, form A, Torrance 1966). The others were tested by situational, behavioural probe measures, specially devised for this study. The devised measures, although lacking in assessment of psychometric properties, were checked for their face validity with the teachers involved.

In order to measure inquiry questions were posted around the classroom, the answers to which required the use of diverse resource materials. The questions asked were different during each testing occasion but were comparable in difficulty level. Answering the questions was presented as an optional task and children were given a week to perform it. Teachers were instructed not to assist children but to draw their attention to the questions each day. The number of answers supplied was recorded as an index to inquiry level. Cooperation was measured by devising a situation which required children to copy certain lines from a book but there were not enough books to go around. Children could either cooperate with others in performing the task, or copy the lines and return the book upon completion. The number of children who "shared" was observed and recorded as an index of cooperation. Yielding to authority was measured by presenting children with a simple comprehension test. When the test was completed, children were shown the same test as answered by their teachers. Some of the answers were clearly wrong. Children were given a second chance to answer the test. The number of children who changed a previously correct response to be like the teacher's, was taken as an index to yielding to authority. Performance in teachers' absence was measured (in School B only) by requiring children to write repeatedly a simple sentence. For the first five minutes the teachers were present in the classroom; for the subsequent five minutes they left the room and for the last five minutes they returned again. Difference between performance in teachers' absence and presence was taken as an index of ability to work without direct supervision.

Because of time limitations, not all measures were administered during each testing occasion. Tables 1 and 2 below show the timetables of testing occasions in Schools A and B respectively. The tables include information on the measuring of variables relating to teaching methods as well as to students' outcomes.

Table 1

TIMETABLE OF TESTING OCCASIONS, SCHOOL A

| Variable | Pre-move | Post 1 (2 mths) | Post 2 (6 mths) | Post 3 (18 mths) |
|---|---|---|---|---|
| Teaching methods: | | | | |
| Teacher Engaging Class as a Whole | x | x | x | |
| Simultaneous Activities Count | x | x | x | |
| Outcome variables: | | | | |
| Maths: M1 | | | x | x |
| M3 | | | x | x |
| M5 | | | | x |
| Reading: R3 | | | x | x |
| R4 | | | | x |
| Inquiry | x | x | x | x |
| Cooperation | x | x | | x |
| Yielding to Authority | | | | x |
| Creativity | | | | x |

## Table 2

### TIMETABLE OF TESTING OCCASIONS, SCHOOL B

| Variable | Stage I doors closed | Stage II doors open | Stage III doors closed |
|---|---|---|---|
| **Teaching methods:** | | | |
| **Teaching Engaging Class as a Whole** | **x** | **x** | **x (experimental classes only)** |
| **Simultaneous Activities Count** | **x** | **x** | **x (experimental classes only)** |
| **Outcome variable:** | | | |
| **Performance in Teachers' Absence** | **x** | **x** | |

**Results and Discussion:**

**The strategy adopted for the analysis of both teaching methods and outcome variables data was a multivariate analysis of planned orthogonal contrasts (Bird 1975). Where variables were measured more than once, a procedure outlined by Poor (1973) for partitioning repeated measures variation was used.**

**Table 3 shows means and standard deviations for teaching methods data in School A.**

## Table 3

### TEACHING METHODS - MEANS AND STANDARD DEVIATIONS, SCHOOL A

| | E1 | | E2 | | C | |
|---|---|---|---|---|---|---|
| | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. |
| T.E.C.A.W. | | | | | | |
| pre-move | 160 | 121 | 322 | 203 | 318 | 197 |
| post-move 1 | 65 | 92 | 169 | 204 | 181 | 242 |
| post-move 2 | 64 | 138 | 42 | 67 | 316 | 239 |
| S.A.C. cur | | | | | | |
| pre-move | 1.3 | 0.4 | 2.0 | 1.5 | 1.0 | 0.7 |
| post-move 1 | 2.4 | 1.3 | 2.2 | 1.2 | 1.9 | 0.8 |
| post-move 2 | 3.5 | 1.5 | 3.3 | 1.6 | 1.5 | 1.0 |
| S.A.C. non-cur | | | | | | |
| pre-move | 2.1 | 2.0 | 3.4 | 2.3 | 3.7 | 1.3 |
| post-move 1 | 6.1 | 1.5 | 5.3 | 1.2 | 3.6 | 1.4 |
| post-move 2 | 4.0 | 1.8 | 3.8 | 1.7 | 3.2 | 1.8 |

**Table 4 shows means and standard deviations for teaching methods data in School B.**

## Table 4

### TEACHING METHODS - MEANS AND STANDARD DEVIATIONS, SCHOOL B

| | E1 | | E2* | | C1 | | C2 | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. |
| T.E.C.A.W. | | | | | | | | |
| stage I | 278 | 171 | 376 | 105 | 305 | 198 | 269 | 149 |
| stage II | 191 | 63 | 83 | 63 | 184 | 70 | 354 | 89 |
| stage III | 211 | 108 | 194 | 120 | - | - | - | - |
| S.A.C. cur | | | | | | | | |
| stage I | 1.2 | 0.4 | 1.0 | 0.7 | 0.8 | 0.4 | 0.6 | 0.5 |
| stage II | 1.5 | 1.3 | 1.1 | 0.3 | 1.0 | 0.0 | 0.9 | 0.5 |
| stage III | 1.1 | 0.3 | 1.4 | 0.5 | - | - | - | - |
| S.A.C. non-cur | | | | | | | | |
| stage I | 3.4 | 1.8 | 4.4 | 1.0 | 3.6 | 1.3 | 4.8 | 1.4 |
| stage II | 3.6 | 1.7 | 3.8 | 2.0 | 3.2 | 1.7 | 3.9 | 1.6 |
| stage III | 3.1 | 1.3 | 2.8 | 1.9 | - | - | - | - |

Table 5 shows means and standard deviations for outcome variables in School A.

Table 5

OUTCOME VARIABLES - MEANS AND STANDARD DEVIATIONS, SCHOOL A

| Variable | Testing occasion | E1 | | E2 | | C | |
|---|---|---|---|---|---|---|---|
| | | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. |
| M1 | post 2 | 19.1 | 8.9 | 17.5 | 6.5 | 27.3 | 7.1 |
| | post 3 | 27.7 | 10.5 | 25.4 | 8.6 | 38.2 | 7.6 |
| M3 | post 2 | 14.0 | 7.2 | 12.5 | 3.9 | 14.2 | 2.8 |
| | post 3 | 21.1 | 6.8 | 18.8 | 7.3 | 25.5 | 6.9 |
| M5 | post 3 | 18.0 | 6.3 | 17.0 | 5.8 | 20.2 | 5.3 |
| R3 | post 2 | 99.5 | 14.4 | 103.4 | 15.7 | 108.5 | 15.1 |
| | post 3 | 108.3 | 14.9 | 107.7 | 17.1 | 115.0 | 16.5 |
| R4 | post 3 | 96.4 | 9.9 | 101.7 | 14.9 | 103.1 | 14.6 |
| inquiry | pre- | 11.7 | 13.0 | 3.1 | 7.2 | 7.6 | 10.0 |
| | post 1 | 3.4 | 6.3 | 4.7 | 10.9 | 2.9 | 7.4 |
| | post 2 | 3.9 | 10.1 | 9.4 | 12.7 | 0.5 | 2.3 |
| | post 3 | 10.0 | 12.8 | 10.5 | 13.9 | 4.1 | 8.4 |
| cooperation | pre- | 7.3 | 4.3 | 7.7 | 4.1 | 8.3 | 3.7 |
| | post 1 | 6.9 | 4.6 | 9.4 | 2.2 | 7.7 | 4.1 |
| | post 3 | 9.5 | 2.0 | 8.8 | 3.1 | 7.7 | 4.1 |
| yielding to authority | post 3 | 11.7 | 9.1 | 15.8 | 7.7 | 11.8 | 9.5 |
| creativity | post 3 | 23.1 | 8.7 | 23.5 | 12.3 | 25.0 | 17.2 |

Table 6 shows means and standard deviations for outcome variable in School B.

Table 6

OUTCOME VARIABLE - MEANS AND STANDARD DEVIATIONS, SCHOOL B

| Performance in teachers' absence | E1 | | E2 | | C1 | | C2 | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. | $\bar{X}$ | S.D. |
| stage I | | | | | | | | |
| Teacher in | 78 | 32 | 66 | 31 | 100 | 27 | 108 | 27 |
| Teacher out | 67 | 44 | 30 | 19 | 75 | 33 | 67 | 54 |
| Teacher in | 90 | 43 | 54 | 34 | 192 | 28 | 92 | 61 |
| stage II | | | | | | | | |
| Teacher in | 125 | 60 | 88 | 32 | 154 | 39 | 154 | 64 |
| Teacher out | 93 | 54 | 56 | 38 | 112 | 41 | 81 | 61 |
| Teacher in | 104 | 53 | 68 | 38 | 136 | 35 | 96 | 55 |

In terms of the operational definition of openness adopted in this research, the experimental classes in Schools A and B could not be regarded as equally open. Observations of organizational strategies revealed that in School A the experimental classes, when in the open space classroom, adopted what are commonly considered open education practices: the contract system was employed, no timetables operated and there were no fixed seats. In the control class, traditional education practices were employed: the lesson system was adopted, timetables operated and children sat at fixed seats. At School B, both experimental and control classes (the former under conventional as well as open space settings) operated according to traditional organizational strategies.

The measurement of teaching methods showed significant differences in School A between the experimental and control classes. Significantly more activities occurred simultaneously during a given observation in the experimental than in the control class. The teachers of the experimental class engaged the class as a whole less than did the teacher of the control class. It is of interest to note that shortly after the move of the experimental classes to the open space classroom (at post 1) the teacher in the control class also shifted her teaching methods towards 'openness': she reduced her whole-class teaching and increased the diversification of simultaneous activities. These changes, however, were temporary. By the end of the year (at post 2) the control class teacher returned to her former conventional teaching methods. The increase in non-curricular activities per observation period at post 1 relative to post 2 probably reflected a 'settling in' period after the move from a conventional to an open space classroom.

At School B a significant change towards openness was detected only in one teaching method: the experimental class teachers reduced their whole-class teaching when in an open space setting. A significant reduction in whole-class teaching occurred also in the control class which moved to the new conventional space classroom.

On the basis of the operational definition of openness and traditionality adopted for this research, the experimental and control classes at School A could be regarded as open and traditional respectively. In School A however, lack of differences in organizational strategies and in the number of activities occurring simultaneously when in an open space setting meant that the experimental groups could not be fully regarded as open classes, even though they operated in an open space environment and experienced a reduction in whole-class teaching.

The difference in the way the two schools responded to the introduction of the innovation may have been due to the circumstances in which it was introduced in them. School A being a demonstration school was probably more 'ready' to integrate the innovation. The teachers in the open space classroom wanted to teach according to the principles of open education. In School B, where teachers were not familiar with open education methods, physical changes in the environment were not accompanied by organizational changes. Discussions with the teachers revealed that they reduced their whole-class teaching in order not to disturb each other.

The battery of product variables administered in School A showed no statistically significant differences between the open and traditional classes on variables which were measured once, or in gains on variables which were measured more than once. Both classes gained significantly in attainment from one year to the next in all cognitive variables, with no significant difference between gains in the open and traditional classes.

The measure 'Performance in Teachers' Absence' in School B showed that the experimental class was less affected in its performance by the teachers' absence, when it was in an open space setting, than when it was in a conventional space setting. Of the two control classes, the one in the new conventional space classroom was less affected by the teacher's absence than was the control class in the old conventional space classroom. This is an intriguing finding in view of the fact that in the control class in the new classroom a reduction occurred in whole-class teaching. It is likely that a relationship existed between amount of whole-class teaching and performance in teachers' absence: the less the teachers engaged the class as a whole, and the more they dealt with individual children, the more independent and autonomous these children were in their approach to work, regardless of the fact that the class was not really 'open'.

An inspection of differences between means revealed a trend (though not significant) for open class children at School A to perform better on behavioural probe measures (inquiry, cooperation and yielding to authority) and for the traditional class children to perform better on standardized paper-and-pencil tests (mathematics, reading and creativity). Informal observations of children during the administration of the behavioural probe measures revealed that open class children displayed a greater degree of independence in their test-taking behaviour. This was noticed particularly during the administration of the 'yielding to authority' measure: children in the open class questioned the researcher about the credibility of the teachers' 'answers' and generally appeared less docile and conforming than the traditional class children.

These trends suggest that standardized paper-and-pencil tests may indeed put open class children at a disadvantage. The devised tests represent a rare effort at measuring the special attributes of open education by measures which require children to perform in a setting more akin to their usual classroom activities than to a test situation. These tests, although checked with teachers for their face validity, were necessarily limited in their scope and quality. Future research should develop behavioural probe measures and subject them to tests of reliability and validity. Children's approach and behaviour during the administration of the measures should be systematically observed to yield relevant and useful information.

## Interview Study, School A

Open classes are purported to be child-centred environments, where children's well-being and viewpoints are of paramount importance. Nevertheless, previous research has largely neglected to examine children's views on open classes. In order to correct this omission interviews were conducted on four different occasions in order to assess children's views of and opinions about the special features of their classes and to examine changes in views over time. It seemed of interest to examine the extent to which the open class exerted influence on other classes in the school. For this reason traditional class children were also interviewed about what for them was a learning situation which they viewed from outside. The questions asked related to the objectives which the designers and teachers hoped to achieve and to educational issues raised by parents. Children were asked about their specific likes and dislikes, their ability to relate to two teachers in a double-sized class, their feelings about the spatial flexibility of the open classroom and their preference for fixed or free seating arrangements.

With regard to general likes and dislikes the impression gained from the interviews was that most children were satisfied with the open class. They gave a variety of reasons for their satisfaction, the most frequently mentioned being the spaciousness of the classroom and its cosiness due to the carpet. The proportion of children who liked their class unequivocally increased with familiarity and experience with the open class situation. Awareness of the special features of the open class also increased with familiarity. The features which were viewed as most attractive just before and shortly after the move were not necessarily connected with open space structures: any class can be equipped with electrical gadgets or with carpet and new furniture. With increased familiarity, however, children tended to mention aspects which coincided with the objectives of the designers of the open space classroom, as sources of satisfaction. Team-teaching, group work, increased choice of activities, flexibility of timetables and of movement were viewed as advantages.

Despite the overall satisfaction, some dissatisfaction was also apparent. A few children viewed critically the noise and crowdedness of the open class. In fact, the interview served to identify five children who consistently made critical remarks. The degree of adjustment of these children to an open class situation is questionable.

As for the specific features of the open class: with increased familiarity children generally became able to identify with both teachers as theirs and clearly perceived the advantages of team teaching. Despite the fact that children were allowed to move about the classroom freely, most children indicated specific spots which were their 'favourite'. Areas which were perceived as quiet and in which "one could concentrate" were particularly liked. These areas, although in no way acoustically treated, tended to be visually cut off from the rest of the class (corners, nooks created by divider units, etc.).

Traditional class children's reaction to the open class was of interest: parents' concern that children who remained in the traditional class will resent and envy the open class, was not justified. These children appeared satisfied with their own class. They were aware of the special attributes of the open class but no envy or resentment were detected. On the basis of the interview data there seemed no reason why open and traditional classes cannot operate side by side within the same school. Some features, such as free seating, particularly liked by the traditional class children could easily be introduced within the framework of traditional class organization.

In an overall appraisal of the interviews, children's articulateness and willingness to provide relevant information about their perceptions and attitudes must be noted. Clearly children can be usefully consulted about their views when innovative educational methods are introduced.

Observation Study, School A

Previous open education research has tended to examine the performance of groups of children. This study extended the evaluation of open education a step further: systematic behavioural observations were conducted in order to examine individual children's styles of behaviour. More specifically, the purpose of the observations was to examine the ways in which individual children varied in the use they made of the open class organizational features, namely, choice of activities, absence of timetables and free movement within the class. The questions asked were: what proportions of time were spent on the various activities? Were some activities more favoured than others? How did children organize their days with regard to the diversity of activities? During the course of one day do some children engage in more different activities than others? How do children vary with regard to lengths of concentration spans in one area or in one activity? Do they shift frequently from one area to another and from one activity to another? What use do children make of the spatial properties of the classroom? Are all classroom areas equally used?

Twelve children were observed in a systematic schedule over a five week period whilst in fourth grade (Roseth 1976). Rather than observe a random sample, the children observed were the six who were rated by their teachers as 'well adjusted' and the six who were rated as 'poorly adjusted' to open class situations.

Results of the observation study showed that children displayed a marked degree of individuality in their classroom behaviour. They behaved according to individual styles in their choice of activities, the amount of time they devoted to the various activities, their perseverance on any activity, and their use of the various classroom areas. Children who were rated by their teachers as 'well adjusted' to open class situations spent more time on curricular and less time on non-curricular activities, than children who were rated as 'poorly adjusted'. However, the degree of perseverance on a given task of the poorly adjusted children was similar to that of the well adjusted.

What is sufficient time spent on curricular activities is a matter of value judgement. In this study, however, children were not found to avoid learning activities. Despite the freedom given to children to organize their day, they were not observed to 'loaf' or 'waste time' for long periods. Eleven of the twelve children spent at least one-third of their time - and some considerably more - on curricular activities. Not all activities were equally pursued, nor all classroom areas uniformally used. Children spent very little time on social studies and science activities; in spite of its spaciousness and equipment, the practical area was little used.

## Conclusions

In an overall appraisal of this research programme care must be taken when generalizing the results obtained from the two schools studied, to other schools. The main finding was that open education does not produce better mathematicians or readers, nor even children whose sense of inquiry, cooperation, conformity and creativity is significantly more developed than those of their traditionally educated peers. If results are viewed in broader terms than those of statistical significance, the studies did show a tendency for children who received open education to show more inquiry and cooperation and less conformity to the teachers' authority. The studies also found that there is no basis for the often publicized view that open education encourages children to waste time and that children in open classes are not disadvantaged in the acquisition of basic skills.

Those responsible for the implementation of educational policies must be aware that if they set out to encourage the continuation of open education, building open space classrooms is not enough. Without ensuring that teachers in open space classrooms are motivated to open up their approach and are trained in the methods of open education, traditional education will be practised within open space environments.

Open and traditional classes can operate side by side within the same school. This maximizes flexibility and choice for teachers to teach and children to learn in the system which best fits their belief systems.

## References

Angus, M. Some problems and issues in evaluating open education. Mimeo. Perth: Western Australia Department of Education, 1973.

Australian Council for Educational Research and Department of Education, N.S.W. New South Wales Primary Evaluation Programme. Mathematics Tests. A.C.E.R., 1972.

Barth, R. S. Open education: assumptions about learning. Educational Philosophy and Theory, 1969, 1, 29-39.

Bird, K. D. Simultaneous contrast testing procedures for multi-variate experiments. Multivariate Behavioural Research, 1975, 10, 343-352.

Bleier, M., Goveman, H., Kunz, N. & Mueller, E. A comparison of yielding to influence in open and traditional classes. Childhood Education, 1972, 49, 45-51.

Burnham, B. Reading and mathematics achievement of grade 3 pupils in open plan and architecturally conventional schools. Studies of open education, no. 10. York County Board of Education, Aurora, Ontario, October 1973.

Bussis, A. M. & Chittenden, E. A. Analysis of an approach to open education. Princeton, New Jersey: Educational Testing Service, 1970.

Clem, P., Ahern, K., Dailey, N., Gay, M. & Scantlebury, M. A comparison of interaction patterns in an open space and a fixed plan school. Man-Environment Systems, 1974, 4, 59-60.

Corlis, C. Curiosity and openness: empirical testing of a basic assumption. Paper presented to the annual meeting of the American Educational Research Association, New Orleans, 1973.

Crandall, A. H. A comparison of reading attitudes and reading achievement among first grade children in open concept and more formal classes. Doctoral dissertation, University of Connecticut, 1973.

Edwards, P. *An interaction network instrument to assess pupil interaction and movement in open area learning situations*. Vancouver Board of School Trustees. British Columbia: Department of Planning and Evaluation, July 1973.

Featherstone, J. *Schools where children learn*. New York: Discus Books, 1971.

Franks, D. D., Marolla, J. & Dillon, S. B. Intrinsic motivation and feeling of competency among students. *Journal of Research & Development on Education*, 1974, *8*, 20-29.

Grapko, M. F. A comparison of open space and traditional classroom structure according to independence measures in children, teachers' awareness of children's personality variables and children's academic progress. Toronto: Ontario Department of Education, April 1973.

Heimgartner, N. L. A comparative study of self-concept: open versus self-contained classrooms. University of Northern Colorado, Greeley Research Study no. 4, 1972.

Hogben, D. Open education: characteristics, definition difficulties, evaluation issues and problems. *The Australian Journal of Education*, 1974, *18*, 225-238.

Innes, R. B. Environmental forces in open and closed classroom settings. *The Journal of Experimental Education*, 1974, *41*, 38-42.

Killough, C. K. An analysis of the longitudinal effects that a non-graded elementary programme, conducted in an open space school had on the cognitive achievement of pupils. Houston, Texas: Bureau of Educational Research and Services, College of Education, University of Houston, 1971.

Klein, Z. & Eshel, J. The active classroom in an integrated school. In N. Degani (Ed.), *Trends in active education*, Booklet 12. Jerusalem: Department of Education and Culture, 1973.

Kohl, H. R. *The open classroom*, New York: Vintage Books, 1969.

Kohler, P. T. A comparison of open and traditional education: conditions that promote self-concept. Paper presented to the American Educational Research Association Annual Meeting, New Orleans, Louisiana, March 1973.

Lueders-Salmon, E. Team teaching and the "active" classroom. A comparative study of the impact of self-contained classrooms and open space team teaching schools on classroom activity. Doctoral dissertation. Stanford University, School of Education, 1971.

McPartland, J. Students reaction to the transition from open elementary school to junior high school: a case study. John Hopkins Univ., Baltimore, Md. Centre for the study of social organization of schools, 1972.

McPartland, J & Epstein, J. L. Social class differences in the effects of open schools on student achievement. John Hopkins Univ., Baltimore, Md. Centre for the study of social organization of schools, 1975.

Moodie, A. G. A survey of reading achievement in a secondary school population. Vancouver Board of School Trustees, British Columbia, March 1971.

Nyquist, E. B. & Hawes, G. R. (Eds.) *Open education*. New York: Bantam Books, 1972.

Owen, S. V. Effects of open education on selected cognitive and affective measures. Paper presented at the annual meeting of the American Educational Research Association. Chicago, 1974.

Poor, D. D. Analysis of variance for repeated measure designs: two approaches. *Psychological Bulletin*, 1973, *80*, 204-209.

Renthrow, R. K., Goldupp, O. & Hurt, M. Development of situational task methodology for the evaluation of process outcomes in the open classroom. Paper presented at the annual meeting of the American Education Research Association, New Orleans, 1973.

Reynolds, R. N. A comparative evaluation of the effects on an open classroom instructional programme and a traditional instructional programme. Harrisburg: Pennsylvania State Department of Education, Bureau of Information Systems, February 1974.

Roseth, N. Evaluation studies of open education in primary schools. Unpublished doctoral thesis, School of Psychology, University of New South Wales, 1976.

Ruedi, J. & West, C. K. Pupil self-concept in an "open" school and in a traditional school. Urbana, Michigan: University of Illinois, 1972.

Sackett, J. W. A comparison of self-concept and achievement of sixth grade students in an open space school, self-contained school and departmentalized school. Doctoral dissertation, University of Iowa, 1971.

Scheirer, M. A. A study of the effects of open classroom education on children's achievement, self-concept and attitudes. Master of Arts thesis. State Univ. of N. Y. at Binghamton, Dept. of Sociology, 1972.

Schonell, F. J. *Backwardness in the basic subjects*. Edinburgh: Oliver and Boyd, 1959.

Sewell, A. F., Dornself, A. W. & Gross, D. Controlled multivariate evaluation of open education: application of a critical model. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., April 1975.

Silberman, C. E. (Ed.) *The open classroom reader*, New York: Vintage Books, 1973.

Stephens, L. *The teacher's guide to open education*. New York: Holt, Rinehart & Winston, 1974.

Sullivan, J. Open - traditional - what is the difference? *Elementary School Journal*, 1974, *74*, 493-500.

Torrance, E. P. *Torrance tests of creative thinking*. Norms-technical manual, research edition; direction manual and scoring guide. Princeton, New Jersey: Personnel Press, 1966.

Traub, R. E., Weiss, J., Fisher, C., Musella, D. & Khan, S. *Openness in schools*. An evaluation of the Wentworth County Roman Catholic Separate School Board schools. Ontario: Educational Evaluation Centre, The Ontario Institute for Studies in Education, 1973.

Tuckman, B. W., Cochran, D. W. & Travers, E. J. Evaluating open classrooms. *Journal of Research and Development in Education*, 1974, *8*, 14-18.

# Symbolic Research in the U.S.A.

## Suggestions for research methodologies into Education.*

J.L. Finkelstein

## Introduction

The qualitative researcher is committed to the use of any and all sociological methods which will better reveal the dynamics and processual nature of the situation under study. The investigation of social institutions through an Interactionist framework can demonstrate the salient characteristics of human exchanges within those locales. With this focus, the researcher is required, firstly, to share an intimate familiarity with the world under study, and secondly, employ a highly flexible methodology. Research skills would be necessary for direct and indirect interviewing; the use of unobtrusive measures including the analysis of body movements, spacing rituals, language styles, exterior body signs or adornments, and so on; the inspection of documents; the construction of life-histories; and the use of participant-observation. Such a combination of methods and data sources is termed triangulation (Denzin 1970).

Essentially, the purpose of qualitative sociological research is to penetrate and understand the reality as lived by its subjects. Under its more common name of Naturalism, this type of sociology is committed to enter (actively) the worlds of individuals and to render them intelligible to others from a theoretical position that is grounded in the behaviours, languages, definitions, attitudes and feelings of those studied. Naturalism attempts to combine covert private features of the social act with its publicly observable and accessible counterpart. Hence, there is a constant movement back and forth between word and deed, intention and action, definition and act, definer and defined. As such, Naturalism moves from simple ethnographic detail to expository and explanatory theory.

The basic unit of analysis in Naturalism is the social act which describes any behaviour where one interactant takes account of at least one other person. It is a symbolic act which can be solitary, joint, covert, public, official or not.

In understanding human conduct, the range and variety of symbols and meanings that are shared, communicated and manipulated by interacting selves in social situations must be encompassed. Society contributes two essentials to concrete interactions: firstly, the symbols or languages which allow communication, and secondly, the settings in which behaviour occurs. Research methods, then, must reveal the symbols used in interaction and the meanings attributed to the social settings in which human conduct is situated. So, before an investigation is complete, symbolic meanings, conceptions of self in social settings, the attributes of social settings, and the nature of processual exchanges must be revealed. Research methods, then, must be capable of capturing the interacting elements of human discourse and of providing concepts that are sensitising rather than operational. By sensitising, it is meant that concepts are deliberately left open until the researcher has entered the field and successfully learned the specific meanings that are socially exchanged. Examples of this grounded process are Goffman's concept of stigma, Laing's characterisation of the family as schizophrenogenic, Farberman's notion of criminogenic (particularly, as used by Denzin in capturing the modus operandi of the liquor industry). In the concluding section of this paper we suggest that a sensitising concept useful for educational research would be to see the school as a conspirogenic situation, and in so doing, understand more of the educational process.

---

*This paper is drawn directly from the on-going research into the social significance of liquor consumption being carried out by Professor N.K. Denzin of the University of Illinois. The application of these research techniques to the sphere of schooling remains my responsibility.

The research techniques of the social scientist are traditionally the experiment, the survey, participant-observation, behaviour specimens, life-histories, unobtrusive measures and the historical method. We shall briefly review these techniques in the light of a Naturalist perspective, thus focusing upon their contribution toward understanding the lived experiences of individuals.

### 1. The Experiment

Crudely defined the experiment is a situation in which the investigator controls some and manipulates other variables thus enabling him/her to observe and analyse the effects of the manipulated variables upon the dependent variables in a situation where other relevant factors are believed to be held constant. However, behaviours within the experimental setting are best viewed as an encounter where the experimenter and the participant are engaged in a form of interaction over and above the formal tasks of the experiment itself. Hence, there are actors (possibly total strangers) in a behaviour situation (the laboratory) that has been given unique cultural meaning. Whilst the ideal experiment flows from a pre-determined script or plan of action, there are other factors going on, concerning the symbolic exchanges between the actors. It is these factors which exist beyond the experiment and which are integral to the encounter between the experimenter and the participant that show the experiment as being a poor means for explaining interactional conduct. As a method itself it ignores the negotiated symbols, the differential meanings held of the setting and the influence of the acting other which are attached to it by virtue of the fact that it is an interactional situation.

### 2. The Survey

The social survey, as a systematic collection of data from populations and samples through the use of the interview or the self-administered questionnaire, has been one of the most popular techniques for sociological investigation. The social survey, however, never proceeds to establish the social basis of human conduct. Respondents are treated as isolated aggregates, only indirectly linked to a social structure. This makes it difficult to establish the basis of symbolic meanings elicited through the interview or questionnaire. Further, establishing equivalence of meaning between the isolated respondents is a problematical, perhaps even, impossible task.

Because of its emphasis upon respondent reconstruction of past events and because of the concentration upon statements rather than episodes of on-going, immediate conduct in the interview, the social survey seems ill-equipped to relate statements to the actual process of constructing meaning - of defining the social situation. It is, however, well suited to charting the contours of stable, simple patterns of interaction. But to continue using it for the examination of complex forms of social behaviour, such as self-concept, attitudes toward social objects, social conceptions of education, the labelling process and so on, can only continue the separation between sociological theory and conceptualising, with empirical sociological findings.

### 3. Participant-Observation

In using participant-observation, the investigator must take the role of the acting other, learn his/her perspectives and study his/her conduct in various settings. Symbols, action and the social nexus of meaning are revealed and related to the daily activities of those being studied. Broadly speaking, participant-observation combines several techniques, namely document analysis, respondent and informant interviewing, direct participation and observation and introspection on the part of the investigator. Document analysis provides historical evidence on events that occurred before the study began and direct participant observation and reflection provide means for relating symbols to the on-going interaction. The method rests firmly upon triangulation and when this is not employed the method can be critised for being anecdotal, impressionistic and even journalistic.

The advantages of the participant-observation technique derive from its suitability for studying complex forms of social interaction, such as social reactions to deviance, face-work, self-emergence, group dynamics, the relations between formal and informal structures within organisations. But without rigour, only unreplicable and generalised statements result. With rigour, the technique is probably the best for entering any level of human interaction, be it on an individual, group or community level.

### 4. Behaviour Specimens

In the practice of naturalistic research, the investigator collects "behaviour specimens" of interactive episodes as these acts arise in their natural settings. The analyst attempts to reproduce in a rich and detailed fashion the experiences, thoughts and language of those studied. The specimens should reflect the actual temporal sequence of behaviour and show how each interactant both influenced and was influenced by all the others in the interactional arena. These specimens, then, can be examined from multiple perspectives - from the points of view of the interactional participants, and from the points of view of social and psychological theory. The investigator compares native accounts of behaviour, of what is going on, with his/her own emerging sociological perspective. Thus, the naturalistic approach blends the temporal, ritual and moral features of the world under study in accord with the sampling and observational strategies of sociology. Measurement procedures are couched in the naturalistic indicators provided from the routinely performed behaviours of the natives in the course of an everyday or week's activities. Such behaviours have a high native relevance and can thus be treated as indicative of theoretical concepts (of a sensitizing nature). Ritual patterns of dress, hair-style and body spacing can be treated as indicators of self-image and relational familiarity. Special languages, codes and dialects reflect group boundaries and relational solidarity. Prized social objects are treated as observable extensions of the person's self, and further, can be seen to indicate prestige, esteem and dignity structures in the subject's world of experience.

The Format of the Behaviour Specimen:

The behaviour specimen will contain:

1. a biographical history of each individual as it is relevant to the encounter at hand
2. provide a physical description of the individual as it is appropriate to the occasion and to the biography
3. a note on any particular speech and behaviour acts, such as stutters, foreign accents, shouting, slurs, whining and so on
4. the specific time at which each actor enters into interaction
5. a description of each self in response to other persons in the arena during the interaction
6. an attempt to reveal what each actor is thinking and ultimately thinks of the behaviour in question
7. a description of social objects which are acted upon in the situation
8. an indication of the salient interactional features of the situation, thus mapping the course on the inter-change and distinguishing a situation from an encounter
9. a projection of the interaction to a conclusion, that is, a suggested outcome.

The behaviour specimen separates the act, the object, the intention, the situation and the interactants so that they can be seen. Like their literary counterpart, the behaviour specimen intends to produce for the reader, the same perceptual and experiential state sensed by the original observer: to give the illusion that the same experiences would have been (and could be) sensed were the reader there. In short, they are reconstructive devices to move the reader to take the role and self of the acting other.

5. The Life History Method

This method represents an application of participant-observation techniques to the single case. Its intention is to obtain a record of the inner, subjective life of individuals and relate it to their on-going relationships and interactions. Life-history materials include documents or records, such as case-histories kept by social agencies; letters; personal communications; biographies and autobiographies; personal diaries and journals.

The flaws of the method show when triangulation is not employed and the investigator has to rely upon only that which is told to him/her. At its best, the technique can reveal generic features and processes that are representative of a totality. For instance, a single life history, when representative, can show a rich picture of behaviour in a broad locale (see Clifford Shaw The Jack-Roller, Chicago. University of Chicago Press, 1966. Also W.F. Whyte Street Corner Society, University of Chicago Press, 1943).

6. Unobtrusive Measures of Observation

The central feature of this technique is the absence of the observer from the social situation under study. The kinds of unobtrusive measures used are:

1. the recording of physical traces left by people - they may be depository (materials left) or erosive traces (the selective wear of materials)
2. the analysis of archival records both public and private, for example, newspapers, public records, literary forms, letters, diaries and so on
3. simple behavioural observations over which the observer has no control, for example, recording dress patterns, gestures, territoriality, body adornments and so on
4. the use of "contrived hardware", namely, cameras and tape recorders in an unobtrusive way.

Whilst unobtrusive measures are effective in mapping out behavioural patterns, they do little in the analysis of symbolic interaction and in the explanation of behaviour. However, coupled with other techniques, unobtrusive measures can be valuable in systematically recording patterns of interaction.

7. The Historical Perspective

Social events each have a history, and as well, are embedded in specific historical contexts that give them their unique meaning and form. So, for analysis of human conduct, a longitudinal-historical perspective is desirable. C. Wright Mills in the Sociological Imagination, especially chapter 8 (New York. Oxford University Press, 1959) is clear on this need, as is Anselm Strauss (Mirrors and Masks, Glencoe, Free Press 1959 , p. 173) and many others.

The value of the historical perspective is that it, firstly, shows the developing, processual nature of the social situation under study, and secondly, outlines the social context, and hence social meanings, that surround the social situation.

The chief sources of data for an historical analysis will be public and private archival documents, as well as the interview. However, in this task, the sociologist differs from the pure historian, in that sociology is more often concerned with cross cultural and cross situational analysis, seeking to build theories of a general kind, whereas the historian can be more directed at detailing specific historical periods and social settings.

## Some Implications of Ethnographic Sociological Analysis

Reality is a social process - an orientation sustained through continuous exchange with others - hence, it is a negotiated definition embedded in an on-going social process. This fundamental assertion of "soft" or qualitative or naturalistic sociology clearly means that there exists no uniformly perceived, empirical world. Furthermore, it means that research itself is a social process with a negotiated existence and meaning. Thus, theoretical, empirical and methodological realities are realities fraught with conflicting and problematical definitions and meanings. As such, the reality of any theory or method is distinct and unique, and to expect validation (as traditionally defined) with any other perspective is an inappropriate expectation. Hence, we recognise that each methodology will lead to a slightly different conception of the world.

The implications of such a position are:

1. that the activities of the scientist/researcher are basically no different from those of the actors in the everyday world which s/he studies. Except that the researcher may be confined by a set of rules established by his/her colleagues that determine the admissibility of data and their transformation into theoretical schemes and concepts:
2. that a triangulated methodology will present unique elements or facets of the same object, and that automatic validation by different research methods cannot be expected. Further, not only will each research method reveal a slightly different aspect of what is presumably the same object, but each investigator who employs that method can be expected to generate slightly different interpretations:
3. that even when the same methods, the same objects and the same investigators are combined there will still be different results, simply because the world is in constant process. Each time we investigate the constructed world different findings and observations will emerge:
4. that theoretical positions and methodological techniques must be flexible enough to grasp this shifting nature of the empirical world. Contingent interactive propositions are needed, not propositions assuming stability and direct causality. Further, theories and research strategies are elements embedded in an on-going social process. Hence, the social reality of the scientist is a political reality - one in which competing forces, institutions, personal and public careers all combine to forge unique definitions and interpretations with distinctly personalised, even ideological elements. Hence, the sources of change in theories and data derive, not only from negative empirical findings, but more from the perspectives of individual researchers involved in the process of reconstructing a world, neither fully their own nor totally foreign. (Denzin 1970a).

## A Suggested Ethnographic Application to Education

In applying the Naturalistic perspective to research into the schooling experience, we can see that the main classes of people involved are the teachers and students who are engaged, principally, within a classroom situation. Beyond this focus there is an extensive service network encompassing the dynamics of the staffroom, through to community influences, teacher-training programmes, Governmental policy and so on. To understand this multi-faceted educational process, we can begin with ethnographic material, much of which has been undertaken in recent years. Such material describes many of the facets of a school;

a. the temporal order.
   The hours in which a school is operating.

b. the cyclical structure.
   The rhythm of interaction within the school at various times of the day; the differential use of school facilities at these times, etc.

c. the social actions and interactions.
   The styles and manners of classrooms, staffrooms, outdoor grounds, and the ways in which self is presented by school-members in various locations and situations.

d. the social types.

A description of school members in terms of conventional sociological concepts, race, sex, class, status, religion, etc.

e. the interactional/person order.
The recording of patterned behaviours within specific locales, such as the regularity with which certain individuals occupy areas of the school. For instance, the head-teacher, student teacher, top student, local bully, potential drop-out etc.

f. the everyday etiquette practices.
The order and disorder of everyday transactions within the school classroom, including the influences of age, sex-role, relational status, and so on, upon transactions; the rules governing visual and tactile contacts, verbal exchanges and emotional expressions.

g. the teaching structure.
Who teaches what to whom and how often.

h. the teaching encounter.
The dynamics of the teaching-learning process including the pedagogical structure of knowledge through to the physical appearance of the classroom, teachers, and students.

An Ethnography produces a complex sociometric picture of life within a school, as well as providing information on (1) the categorisation of the school type, (2) the history of educational advances during a specific time period, (3) the definition of the teacher and student role, and (4) the nature of order and disorder within the school as a social institution. In short, it provides the necessary background from which to select sensitising concepts.

In recent years, research into education has been scrutinising the assumptions of conventional schooling. Teachers and students, themselves, have become more articulate and public with their disappointments and expectations of education. Within such a climate, the nature of the schooling process is becoming more easily identifiable. An in-road into the schooling experience is to employ sensitising concepts in research. One such concept is the idea that schooling is a conspirogenic situation. By conspirogenic, we mean that the definition of the situation is frequently beyond the control of the immediate participants. So, students are regarded as unable to monitor their own educational needs and progress; the standards applied to them originate beyond their sphere of power; and their achievements and abilities, irrespective of quality, are generally regarded as both inadequate and partial. Hence, the student is urged to follow a programme in which the goals and the achievements are frequently beyond his/her influence; for instance, the hidden and survival curricula. In this way, the student is subject to the decisions of others without having recourse to alternatives, and without witness to the rationale for the decisions. This situation not only applies at the level of curriculum design, school policy and the organisation of knowledge, but it also exists withmthe classroom where conventionally the teacher holds absolute authority and there is no collective student power. (Young 1971).

The notion of the school being a conspirogenic situation not only holds for the student, but also involves the teacher. In an organisation, such as the school, where there is a hierarchical and bureaucratic order, only a few individuals have access to decision-making processes. The teacher, like the student, is subject to decisions concerning curricula, assessment, personal appearance and demeanour, etc. in which they are unable to participate. For the student and the teacher, the school can be a situation in which knowledge (be it academic and/or administrative) is mystified, pedagogical and coercive.

The general tone of the criticism against conventional schooling, in recent years, has focused upon the conspirogenic nature of the institution. The ways in which schools have selected, streamed, categorised, and determined the performances, abilities and futures of students has been a central criticism. As well, the ways in which schools appear to survive on the basis of their hidden powers which prohibit their clients from questioning these processes, has constituted a further area of criticism.

The notion of schools being a conspirogenic situation is an example of how sensitising concepts as used within naturalistic sociological research can be developed. The way in which the concept can be applied at the various levels within the schooling process (for instance, the classroom, the staffroom, the teacher training process, Governmental policy-making, etc.) indicates its value as a tool for penetrating the phenomenon of schooling. And, further, its reliance upon a background of extensive research and data collection provides a means for triangulating a great many independent studies that have been carried out over several years.

## References

Denzin, N.K.

1970 *The Research Act in Sociology*.

1970a "The Methodologies of Symbolic Interaction: A Critical Review of Research Techniques", in G. Stone and H. Farberman (eds.) *Social Psychology Through Symbolic Interaction*.

Young, M.F.D.

1971 *Knowledge and Control*.

## SATURDAY 12 NOVEMBER

## 2.00 - 3.00pm

| PAPERS | Rooms |
|---|---|
| WORKSHOP ON CURRICULUM EVALUATION | Ladies Drawing Room |
| J. Biggs, P. Chopra (University of Newcastle)<br>Pupil Evaluation of Student Teachers | Seminar Room A |
| N. Hart, R. Walker (Mt. Gravatt C.A.E.)<br>An Analysis of Child Language in Use<br>(Part 1: N. Hart; Part 2: R. Walker) | Seminar Room B |
| K. Sinclair (University of Sydney)<br>Generational Differences in Adolescent Intellectual Development | Gateway Room |
| W.N. Bardsley (W.A.I.T.)<br>Assessment of School Environments | Meeting Room |

# Pupil Evaluation of Teachers

John Biggs and Prem Chopra

University of Newcastle

## Abstract

While student evaluation of tertiary teaching is by now well established, there has been little work reported on the use of pupil feedback as formative evaluation for secondary teachers. The present investigation is concerned with (a) constructing a pupil evaluation of teachers (PET) scale, using dimensions of teaching behaviour and of affective pupil responses to teachers, that can be assessed reliably and validly throughout years 7 to 11; and (b) using the scale as a source of feedback to both regular and student teachers for improving teaching behaviour.

After a series of studies, PET was reduced to four areas of teaching behaviour -- Management, Negative Focusing, Faulty Verbalization and Teaching Clarity -- and one overall scale of Pupil Reaction to the teacher. Internal consistency and test-retest reliability were high in all scales for regular and student teachers, while between-class agreement was high for regular but somewhat lower for student teachers. Ratings were little affected by sex of pupil or teaching year of pupil or ability level.

PET was used in two feedback studies involving 19 regular and 23 student teachers but no evidence was obtained in either case that teaching behaviour had changed in a way perceptible to pupil raters.

# PUPIL EVALUATION OF TEACHERS[1]

John Biggs and Prem Chopra
University of Newcastle

## Introduction

Student evaluation of teaching has long been established at the tertiary level as a valid and reliable means of assessing teaching performance (Costin, Greenough & Menges, 1971; Doyle, 1975; Miller, 1972). Amongst the many uses to which student evaluations have been put, a particularly fruitful one is that of providing feedback to teachers so that they can correct faults in their presentation, course design, etc. (Trent & Cohen, 1973). In a recent study in this country, Oie and Chopra (1975) showed that when university lecturers in a first year Biology course were provided with feedback from students, in a retest students perceived the lecturers' performance as having changed in a desirable direction. Furthermore, it was shown that this change was perceived by students in different classes than those carrying out the original rating.

For many years now, teacher educators have been using feedback from many sources to provide both formative and summative evaluation for student teachers during practice teaching. Sources of such feedback derive from their own observations of method lecturers and those of teachers, particularly "master teachers", in the practice teaching schools; and of course, in recent years there has been a great deal of use of audio, and audio-visual, sources of feedback in both macro and micro teaching situations.

There appears to have been little attempt made to utilize the pupils' perceptions of the teacher as a source of useful feedback. (Note the term "pupil" will refer below to school students, and "student", or "student teacher", to tertiary students). This would seem to be unfortunate since it is quite likely that pupils would have better grounds for evaluating student teachers' performance than other observers, in that

1. Observers, even master teachers, do not have nearly as long an exposure to the student teachers' performance as do the pupils taught; neither do they see him in such a variety of situations as do the latter.
2. Pupils, as the direct "clients" of the student teacher, bear a relationship to the teacher which is exactly analogous to that of tertiary students; they are exposed to many different teachers, both regular and student, and might thus form a baseline of what to them constitutes good teaching.
3. Given suitable instrumentation, the cost benefits of obtaining pupil evaluations are highly favourable.

There are, of course, possible difficulties with the use of pupil evaluations. The first is to do with the question of whether school pupils have the maturity (to call it that) to make valid and reliable judgments. Second, whether the period of practice teaching is sufficiently long, or sufficiently typical, of a student teacher's actual, and potential, teaching performance to permit a valid evaluation (although this of course applies as well to evaluations by other observers as by pupils). And third, the question of pupil evaluation might well raise non-academic issues, such as the acceptance of the procedure and all that it implies by regular teachers.

In the present investigations, it is attempted:

1. to design an instrument that can be handled by high school pupils i.e., in years 7 - 11.
2. to determine the validity of pupil ratings, both of regular and of student teachers,
3. to determine the extent to which pupils perceive regular and student teachers' behaviours as changing in a positive direction, and if pupils react more positively to teachers who have been provided with feedback,
4. to establish if any such effects generalise across classes,
5. to see if pupil evaluations are related to other variables, such as the sex of the student teacher and the sex of the pupil-rater, the year and stream of the pupils, and the size of the class taught.

The present investigation was conducted in two main stages:

1. The construction of a rating instrument;
2. Investigation of the effects of feedback, both with regular and with student teachers.

## I: Construction of the Pupil Evaluation of Teachers (PET) Instrument

### 1. Selection of Scale Format and Item Types

It was decided from the outset to have two kinds of scale: items referring to categories of teaching behaviour that are overt, and in theory, easily observable; and items referring to the pupils' overall affective reactions to the teacher.

---

[1] This project was funded by ERDC.

### (i) Teaching behaviours

Meighan (1974) refers to four areas of performance along which high school pupils appear to be competent judges:

- (a) Preparation of lessons and interest of lessons
- (b) Certain aspects of presentation, e.g. use of voice
- (c) Attitudes to pupils, e.g. fairness
- (d) General organisation of lessons.

These areas were taken, in the first instance, to be possibly useful ones. They were supplemented by others which curriculum and method lecturers agreed would be useful ones for evaluation purposes, such as questioning skills, classroom management, and flexibility of approach.

The first draft of the questionnaire was constructed around seven such areas, using item types that appeared to the writers, and to method lecturers, as exemplifying that particular area of teaching behaviour. The seven areas were: Classroom Management, Negative Focusing (picking on negative rather than positive pupil characteristics), Pupil Supportiveness (democratic procedures), Faulty verbalization, Interest in Teaching, Teaching for Understanding, Well prepared for class. The questionnaires were administered to pupils in 14 classes in two schools in years 7 and 11 and were factor analysed to a seven factor varimax solution. Each of the items was rated on a five point Likert scale, ranging from "this is always, or almost always, true of this teacher", to "this is never, or hardly ever true of this teacher". After the factor analyses, items with poor loadings were dropped and the scale scores were computed on the remaining items. The original questionnaire contained two items that provided for open-end comment by pupils, and these were given back to the participating teachers, plus a calculation of the means for all fourteen classes, so that individuals could see where they stood in relation to the average of the group tested.

### (ii) Pupil reactions

The pupil reaction set of scales assessed the pupil's personal reaction to the teacher along several dimensions; an overall Good Teacher, Negative feelings about schooling, Self-concept of pupil, Tendency to misbehave, Achievement motivation and reactions to the Teacher as a Person. The aim here was to determine whether such reactions would form a reliable rating scale, and if so, to use the resulting scales as dependent variables, to see if the pupil felt better about the student teacher who received feedback, independently of any perceived changes in the teachers actual behaviour.

The next stage involved recasting the questionnaire in the light of the previous analyses and qualitative comments of both teachers and pupils' comments. In this version, there were 42 items covering the seven teacher performance areas, and 30 items covering six areas of pupil reaction, as described above.

This version of the questionnaire was administered to 36 classes in three Catholic schools. Teachers, when administering the scale, were asked to keep note of items that prompted pupil questioning as to wording or ambiguity of meaning. It was found in this study that there was considerable intercorrelation amongst scales (see Table 4(b)), and in view also of the time taken to administer the questionnaire, it was decided in a third investigation to reduce the scales from seven to five areas of teaching behaviour: Teaching for understanding, Negative focusing, Classroom management, Faulty vocalisation and Teaching clarity. This last dimension had not been included before, but recent work by Cruickshank and associates (Bush, Cruickshank, Kennedy & Meyers, 1976; Bush, Cruickshank & Kennedy, 1977) indicated that this is an important dimension of teaching behaviour, particularly as perceived by pupils.

Similarly, in this third study, the six scales of pupil reaction were reduced to three: good teaching, emotional security and achievement motivation. In addition, a true, ?, false format was tried instead of the Likert format. This version was administered to four classes taught by regular teachers, in a high school. It was found that the clarity and teaching for understanding scales overlapped and so the final version included only four scales of teaching behaviour and only one scale of pupil reaction, while it was decided to revert to the Likert format. The final version of the scale was the one used with student teachers.

## Properties of Final PET Scale

### 1. Reliability and inter- and intra-class consistency

The internal consistency (Cronbach's Alpha), test-retest reliabilities and inter-class correlations (i.e. two classes rating the same teacher) are given in Table 1.

---

Table 1 goes here

---

The earlier scales with regular teachers and the final versions with student teachers all showed a high degree of internal consistency, especially in the case of the Pupil Reaction scale, the only exception being the Management scale with regular teachers. The test-retest correlations were also satisfactorily high in both groups. A pronounced difference between regular and student teachers does, however, occur with the interclass correlations which indicate the extent of agreement between two separate classes in their mean ratings of each teacher. Apart from Management, classes showed a high degree of agreement on all the remaining scales in the case of regular teachers, but in the case of student teachers, agreement between classes was significant only on the negative, and perhaps more strikingly obvious aspects of their teaching, such as Faulty Verbalisation and Negative Focusing. It would appear that the more personal aspects of teaching behaviour show a higher degree of stability than what might be called relational aspects, such as discipline, which depends upon the relationship struck between the teacher and the class. It should be stressed that the interclass correlations for student teachers are based only on a period of exposure of three weeks. Unfortunately, there were only six pairs of classes who were allocated to the same teacher in both block practices, and the correlations between these six pairs over the six week period were all nonsignificant. However, it is fairly clear that length of exposure would have much to do with inter-class agreement.

### 2. Validity

The two curriculum and method lecturers mainly concerned were asked to rate the student teacher, on the same forms as the pupils had used, after each of the two three-week practice teaching blocks. These scores were correlated separately for each supervisor with the class means on all five scales. Numbers were very small and in only one scale (Classroom Management) was a significant correlation obtained between one supervisor's rating and the class rating after the first three weeks. The correlations after a second block of three weeks, were obtained in the case of one curriculum and method lecturer only, the other unfortunately was ill and could not complete this stage of ratings.

---

Tables 2(a) & (b) go here

---

This time, only the Negative Focusing scale showed a significant correlation between the class and the supervising lecturer. While these results might seem disappointing, when the ratings of the lecturer who completed ratings of both practice teaching blocks were correlated, all five were found to be significant. Thus, it seems that both classes and lecturer were reasonably consistent in their ratings of the student teachers but they did not agree with each other. In view of the individual consistency and stability of both class and lecturer, it thus seems that their respective viewpoints of the student teachers were complementary: each saw different, but relatively constant, aspects of the student teacher's performance.

### 3. Factorial Structure

The final scale was factor analyzed. The data are reproduced in Table 3.

---

Table 3 goes here

---

It can be seen that the biggest factor was defined by the ten pupil reaction items and all loadings were highly significant. Most other scales came out fairly cleanly with the appropriate scale item loading significantly, but there was some overlap, particularly on the Negative Focusing scale which tended to draw items from the Faulty Verbalisation factor. Some of this overlap is no doubt due to wording of items: (e.g. "shouts and yells too much" was meant as a Negative Focusing item but it also says something about Faulty Verbalisation). Logically, however, these are two different dimensions of student teacher performance, and so both were retained as separate scales. The intercorrelations between the scales are shown in Table 4.

---

Table 4(a) & (b) go here

---

When class means are calculated and intercorrelated it can be seen that there is some overlap between scales, as suggested by the factor structure. Student teachers who can manage their classes tend also to be rated as clear in presentation. Students who focus on negative aspects of pupils tend to be rated as poor at verbalisation, unclear, and are reacted to unfavourably. Those with Faulty Verbalisation are similarly rated as unclear, and as potentially "poor" teachers; while those who are rated as clear in their teaching, receive favourable pupil reactions.

The equivalent correlations are given in Table 4(b) for the regular teachers rated on the larger, earlier version of the instrument. As can be seen, the intercorrelation of scales here is considerable: Pupil support, Interest, Understanding (the earlier version of Clarity), Preparedness, and Good Teacher, on the one hand, versus Negative Focusing and Faulty Verbalisation on the other. Pupil reaction is neutral towards good (or poor) classroom managers, but is very closely related to their perception of the teacher's interest in them. The final version of the scale does appear to offer a more articulated picture.

4. Effects of year, stream and sex of pupil evaluations

In order to use pupil evaluations, whether of regular or student teachers, it is important to find out the extent to which they may be influenced by the ability level, year, and sex, both of pupil and of teacher. Veldman and Peck (1969) showed that grade of class and sex both had effects on pupil evaluations of student teachers in certain areas: pupils rated female student teachers as more friendly than males, and higher grades saw teachers as more friendly, but less interesting, than lower grades. Veldman and Peck used only one scale (Control) that overlapped any of the PET scales and this showed no sex or year effects. In the study on regular teachers, the schools streamed by ability at all year levels. Accordingly, a 3 x 2 ANOVA was carried out on all the teacher behaviour and pupil reaction scales for years 8, 9 and 10 and high vs. low ability streams. There were no significant mean effects or interactions on any of the scales. There was a consistent tendency for year 8, the lowest year, to rate more unfavourably on all scales, but this only reached the .10 level of significance for one of the pupil reaction scales.

In the second study, all the schools were co-educational and there was a mixture of male and female teachers. Accordingly, the class means on each scale were calculated separately for each sex of pupil, and a 2 x 2 ANOVA was carried out on all five scales. There was only one significant main effect and no significant interactions. This showed that on the teacher clarity scale, girls rated teachers regardless of teachers' sex as Clearer than boys ($P < .02$). In the second study, size of class and year of class were correlated with all five scales but, as before, there were few consistent relationships: younger pupils tended to perceive more negative focusing ($\underline{r} = .59$, $P < .01$) as did pupils in small classes ($\underline{r} = -.54$, $P < .01$).

These essentially negative findings are of course very convenient, as they indicate that allowances do not have to be made, within the high school range, for age, stream or sex of rater or of teacher. It was not possible to investigate subject taught, as did Veldman and Peck, as all but two of the student teachers were in the Maths-Science area.

Summary

The above data were gathered from several pilot samples and two fairly large scale studies on the PET questionnaire. It seems fairly clear that it is possible for pupils from years 7 upwards to make stable and internally consistent evaluations of their teachers' behaviours in several areas and to evaluate their own affective reactions to the teacher on at least one major dimension. The dimensions of teacher behaviour that clearly emerged were Management, Negative Focusing, Faulty Verbalisation and Teaching Clarity. Such valuations showed, in the case of regular teachers, a surprising stability over a six week period, and also a relatively high degree of agreement between two classes taught by the same teacher. The ratings of student teachers however were nearly as stable over the same class, but not so consistent across classes.

There was very slight agreement between the student teachers' curriculum and method lecturers ratings and those of the classes taught. The reason for this could either reflect on the validity of pupil ratings, (but not reliability, in view of other data), or could simply mean that the pupils are rating validly, but on different dimensions to the curriculum lecturers, each rating their different perspectives and exposures to the teacher.

The influences of year, stream, and sex of both pupil and teacher on the PET scales were shown to be slight.

## II: Feedback Studies

Two studies of the effects of pupil feedback on student performance were carried out. The first was with regular teachers and the second with student teachers.

### 1. Regular Teachers

The principal and staff of three Catholic schools agreed to participate in the feedback study. There were 19 teachers involved, completed data being available for a total of 36 classes. All were experienced, including several religious teachers, and the nature of the experimental design was explained to them. Teachers were allocated to feedback and non-feedback groups at random. The teachers in the feedback group administered the questionnaires themselves to two classes, while those in the non-feedback group were administered by the writers, or by colleagues, likewise to two classes each, and in neither case were the questionnaires seen by the non-feedback teachers. All data were immediately collected after testing and class means were calculated for each class on each of the seven teaching behaviour and six pupil reaction scales. In the interests of communication, the raw scores were standardised to a mean of 100 and a standard deviation of 15, the total population being the class means of all classes tested, so that the teacher by looking at a particular scale score and, using the analogy of IQ score, could tell immediately whether he was above or below the mean for that particular scale for the 36 classes tested and by how much. The original questionnaires and the standardised scale scores with a description of each scale were returned to members of the feedback group only. It was suggested to the feedback teachers that either one of the writers, a curriculum and method lecturer in their own area, or a colleague, could go through their questionnaires with them and perhaps observe them teaching in a classroom, and in this way try to reinforce the messages given by the questionnaires. None of the feedback group accepted this offer and they perused the questionnaires and the processed material on their own. Six weeks after the first testing, the entire group were re-tested, and the same calculations for each scale carried out.

The effect of feedback would be assumed if the pupils either perceived the teachers' behaviours as having changed, and/or if they reacted more favourably to the teacher after feedback, as compared to the non-feedback controls. The data fitted a three-way repeated measures ANOVA; Class 1/Class 2 x Feedback/Non-feedback x Pretest/Posttest, the last being a repeated measure; on each of the 13 scale raw scores.

However, none of the main effects or interactions were significant and the data are not reproduced here. There was, then, no evidence that feedback enhanced the pupil perception of regular teachers.

### 2. Student Teachers

A major point of the investigation, however, was to discover if feedback could be usefully employed for improving student teachers' teaching behaviours. It would be hoped that these behaviours would still be sufficiently malleable for such improvement to take place, particularly if bolstered by one to one interaction between student teacher and his curriculum lecturer. Accordingly, arrangements were made with three curriculum and method lecturers who were willing to participate and who obtained volunteer student teachers from their classes. There were 12 students in the feedback group and 11 in the non-feedback. Each student was asked to test two classes who knew him well. The arrangements for practice teaching were that in second term, three weeks practice was followed by four weeks back at University, with particular reference to analyses of school experience, followed by a second three weeks in practice teaching. The pre-test was to be done during the last week of the first block, and the post-test, in the last week of the second block of practice teaching. Immediately after testing, the subject master took charge of the questionnaires and they were picked up by a member of the University staff before any student teachers were able to see them.

There was a total of 43 classes available for the pre-test, and the means of each scale were calculated as before, and standardised to a mean of 100, with a standard deviation of 15. This work was completed at the end of one week and during the second week of the intermission, the original questionnaires, and the analyses by class, were returned to the feedback students, who were invited to consult with their curriculum and method lecturer for the remainder of the period. Most students did so, and attention was particularly drawn to the information from the questionnaires that could help them improve particular items of teaching performance. They would then have two weeks in practising the modified behaviour before the post-test was carried out. The non-feedback students did not see the questionnaires; they only interacted with their lecturer in the normal way, without reference to information deriving from pupils, before resuming the second block. Due to logistics, it was found that some students in the post-test were teaching different classes. This made it possible to investigate the extent of generalisation effects, i.e., whether information deriving from one class could be used to change teaching behaviours perceptibly in another class. Accordingly, a 2 x 2 x 2 ANOVA was set up incorporating Feedback/Non-feedback x Same/Different class on retest x Pre-test/Post-test, with the five PET scales raw scores dependent. Only one main effect or interaction was significant; Teaching Clarity declined at the post test ($F = 4.57$; 1,30; $P < .05$). No feedback effects were significant.

## Discussion

There were two major aims of this investigation:

1. To construct a rating scale (PET) of teaching behaviour and pupil reaction that could be used by pupils in the secondary school concerning regular or student teachers; and
2. To evaluate the effects on perceived teaching behaviour of pupil feedback, derived from PET, for both regular and student teachers.

### 1. Construction and Properties of PET

It was found, contrary to opinions expressed by some regular teachers, that pupils were capable of handling this task given that sufficient care was taken over wording of items. Test-retest and internal consistency data showed that pupils were capable of giving reliable ratings, particularly of regular teachers with whom they had had some significant contact. Earlier versions of the instrument were rather too long and the scales overlapped. In the final version, four areas of teacher performance were investigated with eight items for each area and ten items to assess pupil reaction to teachers.

Reliability was assessed in terms of test-retest for the same class, internal consistency, and by correlating across classes taught by the same teacher. Such correlations were highly significant when PET was used for rating regular teachers, but cross-class correlations, although generally significant, were somewhat lower for student teachers, as might be expected in view of their much briefer interaction with their pupils, and of their inexperience and likely instability of their teaching behaviours.

Validity was assessed by correlating supervisors' evaluations using PET with class means, after three weeks and then after six weeks practice teaching. There was little relationship after the shorter, and only a slightly better relationship after the longer, period. As the supervisors' ratings themselves were highly consistent, it seems possible that pupils may well perceive a different aspect of student teachers' performance than do lecturers. If so, important questions might then be raised as to which source, or both, is the most useful (a) for formative, and (b) for summative evaluation. The last question, on summative evaluation, is the most easily answered: there are practical and industrial reasons why pupil evaluations cannot be used for grading students, not the least being that this was one of the conditions laid down by the NSW Teachers' Federation. Clearly, more research is needed on this question of what are valid and useful viewpoints of student teaching performance. One viewpoint we did not investigate, for example, was that of the school-based supervisor or master teacher.

Intercorrelations between scales were very high in earlier versions of PET; only moderately so in the final version. The latter's factor structure fitted the final assumption that items could be construed in four dimensions of teacher behaviour and one of pupil reaction. While Management was unrelated to most other factors, except Clarity, Negative Focusing (e.g. sarcasm) was strongly related to unfavourable pupil reaction, while Clarity seemed to produce good pupil reaction.

The influence of several factors on PET subscale scores were investigated: year of pupil raters, their ability level as assessed by stream, sex of pupil, sex of teacher and their interaction. Such influences were found to be slight. There was some evidence that younger pupils tended to perceive more negative criticism, and also there was a higher degree of negative criticism perceived in smaller classes.

2. Feedback studies

There was no evidence that the provision of feedback to teachers and student teachers has any significant effect on the pupils' perception of teaching behaviour.

It is suggested that the lack of significant improvement as a result of feedback is a result of one or more of the following:

(i) The instrument, PET, is not sensitive enough to record changes that may have occurred. On the other hand, PET scales did show a high degree of reliability and factorial stability. It is therefore possible that, while we did try and focus upon important and observable aspects of teacher behaviour, the nature of the feedback provided by the PET scales is not that which is of value in changing pupil perceptions of teaching behaviour. Further research into the question of feedback from pupils, with and without PET, is clearly desirable.

(ii) The provision of feedback had insufficient impact. This was undoubtedly true in the study using regular teachers, as none of those concerned accepted the invitation to use a third party as consultant and counsellor, so that the pupil evaluations could be discussed and supported as appropriate by the observations of the third party. This objection has less force in the case of the student teachers, since most of the feedback group did voluntarily discuss the pupil feedback with the curriculum lecturer. In these discussions, special consideration was given to those scale items which were clearly below average, and both the lecturer and the students expressed surprise when the ratings by the pupils showed no difference in these behaviours (Negative Focusing was a particular case in point with one student). It is possible then that where the teacher has the same class, the pupils' evaluations are, as it were, "frozen" by their first interaction with the teacher, and they tend therefore to see him in that light even if, in fact, he has changed since. This should not have been the case when the student taught different classes, but as noted, there was no Feedback x Different class interaction. On the other hand, if the different class did in fact see a different aspect of the teacher, as other evidence suggests, then this Feedback x Different class interaction would have in any case been washed out.

(iii) Teaching habits are well established and not liable to be easily changed. The results of the study with regular teachers are most compatible with this suggestion in that the ratings showed a high degree of stability and no susceptibility to change. In the case of the student teachers, this argument is harder to sustain in that teaching behaviours did appear to be unstable and presumably therefore the more easily influenced. The fact that they were not so influenced suggests that there was insufficient impact, or "digestion", of the pupil information, or that the information as provided was not the most relevant for producing change.

(iv) It is possible that change was produced, but that the pupils (who have already been noted as operating on a different dimension that supervisors) rated on aspects of teaching performance that were unrelated to the feedback provided. This would seem to be highly unlikely, however, because the feedback was precisely on those dimensions on which the pupils subsequently rated.

(v) The teachers in various ways "defended" against those aspects of the pupil ratings that they did not want to hear. In the case of regular teachers this could be done fairly easily, simply by regarding pupils' evaluations as a good P.R. exercise, but as essentially irrelevant to their teaching skills. By the same token, since the pupil evaluations could not be used for summative evaluation, it could be argued that the students would similarly devalue them and see them as irrelevant to improving their teaching performance. It is difficult to assess how real these objections are. Both regular and student teachers were volunteers, and it would be inconsistent for them to go to the trouble of cooperating in the study, and then by ignoring the feedback render the whole exercise pointless.

(vi) The physical arrangements, in particular the time intervals between the provision of feedback and post-testing (six weeks in one study, four weeks in the other) may be critical.

In summary, while the feedback studies were not successful, there are several possible reasons for this. Clearly further research needs to be conducted to establish the optimum conditions under which feedback may be used to enhance or remedy teaching performance, both of student teachers and of regular teachers on an inservice basis. Such research will need to look at the mode of presentation of the feedback and especially the effects of active interpretation of scores by "experts", and the use of remediation exercises. There is a strong prima facie case for using pupil feedback, both in that it would have an obvious face validity and in that the cost benefits are most favourable.

## References

Bush, A.J. Kennedy, J.J. & Cruickshank, D.R. An empirical investigation of teacher clarity. Journal of Teacher Education. 1977, 28(2), 53-58.

Costin, F., Greenough, W.T., & Menges, R.J. Student ratings of college teaching: reliability, validity and usefulness. Review of Education Research, 1971, 41, 511-535.

Cruickshank, D., Kennedy, J.J., Myers, B., & Bush, A.J. Teacher clarity - what is it? Mimeographed, Macquarie University, N.S.W., October, 1976.

Doyle, K.O. Student Evaluation of Instruction. Lexington: Heath, 1975.

Meighan, R. Children's judgements of the teaching performance of student teachers. Educational Review, 1974, 27, 52-60.

Miller, R.I. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.

Oie, T.P.S. & Chopra, P. Student feedback and teacher behaviour. The Australian University, 1975, 13, 148-159.

Trent, J.W. & Cohen, A.M. Research on teaching in higher education. In R.M.W. Travers (Ed.) Second handbook of research in teaching. Chicago: Rand McNally, 1975.

Veldman, D.J. & Peck, R.E. Influences on pupil evaluations of student teachers. Journal of Educational Psychology, 1969, 60, 103-108.

## Table 1

### Inter- and Intra-class Consistency in Pupil Evaluations of Regular and Student Teachers

| | Regular teachers (N = 18) | | | Student teachers (N = 23) | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 |
| | Alpha | Test-retest | Interclass | Alpha | Test-retest | Interclass |
| *Management*[a] | .48 | .76** | .48* | .63 | .71** | .42? |
| *Negative focusing* | .64 | .82** | .72** | .74 | .75** | .46* |
| Pupil Support | .70 | .72** | .62** | - | - | - |
| *Faulty Verbalisation* | .56 | .73** | .82** | .71 | .65** | .47* |
| Interest in Teaching | .77 | .80** | .77* | - | - | - |
| *Teach Understand (Clarity)* | .67 | .76** | .65** | .69 | .84** | .41? |
| Preparedness | .66 | .88* | .70** | - | - | - |
| *Pupil Reaction* (overall) | .94 | .65** | .76** | .93 | .72** | .33 |

? P < .10; *P < .05; **P < .01.

[a] Italicized scales remain in the final version of PET. (Understanding similar to Clarity)

## Table 2(a)

### Curriculum and Method Supervisors' Ratings with Class Means

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| Block 1. | | | Man. | Neg. | Verb. | Clar. | Reac. | |
| | Sup. 1 | Class 1 | .52 | .21 | .21 | .39 | .08 | (N = 9) |
| | | Class 2 | .09 | .06 | -.11 | -.36 | -.08 | (N = 9) |
| | Sup. 2 | Class 1 | .83* | -.21 | .48 | .07 | .44 | (N = 8) |
| | | Class 2 | .56 | .10 | .08 | -.20 | .46 | (N = 8) |
| Block 2. | | | | | | | | |
| | Sup. 1 | Class 1 | .23 | .34 | -.09 | -.33 | -.63? | (N = 9) |
| | | Class 2 | -.22 | .71* | .18 | .01 | .18 | (N = 9) |

## Table 2(b)

### One Supervisor's Ratings, Block 1 with Block 2

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Block 1 with Block 2 | .80* | .76* | .79* | .86** | .87** | (N = 9) |

? P < .10; * P < .05; ** P < .01

## Table 3

### Factor Structure of Final Version of PET (N = 1110)

| | I | II | III | IV | V | r with own scale |
|---|---|---|---|---|---|---|
| 1. | .36 | | | .38 | | .57 |
| 2. | | .62 | | | | .57 |
| 3. | | .44 | | | | .56 |
| 4. | | | | .43 | | .54 |
| 5. | .59 | | | | | .59 |
| 6. | | .61 | | | | .62 |
| 7. | | .56 | | | | .53 |
| 8. | | | | .42 | | .55 |
| 9. | .65 | | | | | .52 |
| 10. | | .49 | | | | .61 |
| 11. | | .41 | | | | .53 |
| 12. | | | | .44 | | .57 |
| 13. | | | | .51 | | .55 |
| 14. | | .66 | | | | .68 |
| 15. | | .45 | | | | .59 |
| 16. | | | | .44 | | .55 |
| 17. | .51 | | | | | .60 |
| 18. | | .51 | | | | .63 |
| 19. | | | .52 | | | .66 |
| 20. | | | | .59 | | .67 |
| 21. | | | | | | .39 |
| 22. | | | | | | .45 |
| 23. | | | .63 | | | .52 |
| 24. | | | | .59 | | .60 |
| 25. | .54 | | | | | .54 |
| 26. | | .36 | | | | .55 |
| 27. | | | .66 | | | .62 |
| 28. | | | | .57 | | .52 |
| 29. | | | | .60 | | .52 |
| 30. | | .52 | | | | .67 |
| 31. | | | .58 | | | .58 |
| 32. | | | | .56 | | .53 |
| 33-42 | | | | | .70-.80 | |
| | 5% | 9% | 5% | 8% | 15% | (42% of total extracted by 5 roots) |

Note: Items 1 - 32 constituted the 4 teacher behaviour scales;

Items 33-42 the pupil reaction scale.

### Table 4(a)

### Intercorrelations Between the 5 PET Scales

### (N = 43, class means, Student teachers)

| | | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| A: | 1. | Management | - | | | |
| | 2. | Negative Focusing | .06 | - | | |
| | 3. | Faulty Verbalisation | -.25? | .62** | - | |
| | 4. | Teaching Clarity | .53** | -.43** | -.59** | - |
| B: | 5. | Pupil Reaction[1] | .07 | -.62** | -.51** | .63** |

? $P < .10$; ** $P < .01$.

[1] Rescored so that high means a positive reaction.

### Table 4(b)

### Intercorrelations Between the 8 PET Scales

### (N = 31, class means, Regular teachers)

| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| A: | 1. | Management | - | | | | | | |
| | 2. | Negative Focusing | -.14 | - | | | | | |
| | 3. | Pupil | .45** | -.53** | - | | | | |
| | 4. | Faulty Verbalisation | -.31 | .74** | -.66* | - | | | |
| | 5. | Teach. Interest | .48** | -.70** | .84** | -.74** | - | | |
| | 6. | Teach. Understanding (Clarity) | .38* | -.45** | .88** | -.50** | .82** | - | |
| | 7. | Preparedness | .55** | -.54** | .80** | -.53** | .89** | .82** | - |
| B: | 8. | Pupil Reaction | .34 | -.72** | .79** | -.80** | .91** | .74** | .79** |

$P < .05$; $P < .01$

# AN ANALYSIS OF CHILD LANGUAGE IN USE - (PART I)

(Paper presented by N.W.M. Hart, Mount Gravatt C.A.E., at the
A.A.R.E. Conference, Canberra. November, 1977)

Reading and writing are two facets of language competence which usually develop later than the other two facets, speaking and listening in first language learning. It seems logical then that when children come to learn about written language they should be encouraged to transfer and modify those strategies which prove so successful while they are mastering the oral code.

The work of Goodman et al (1976) and Smith (1971) in America has highlighted the extent to which the reader is dependent on his already developed language skills in the reading process. They have demonstrated that unless the reader can make use of cue systems basic to language processing, reading in the sense of a process for taking meaning from print is impossible.

In addition to recognising that the child's knowledge of language is a necessary component of the reading process, any attempt to make use of this knowledge in teaching reading or language needs to recognise just as urgently that for language to be functional for the child it should not be isolated from his communication with his world. Language should be studied and developed as an interrelated three dimensional system which takes into account syntax, semantics and pragmatics.

Such an orientation was not apparent in the writings of the sixties when we first began our work. At that time there was still stress on the structural, syntactic approaches to the teaching of language and reading.

During the 70's however, the research writings of Cazden (1977), Wells (1974), Halliday (1972), Bowerman (in press), Goodman and Goodman (1976) and many others have succeeded in putting forward research data and convincing arguments that those aiming at using and developing what the child knows about language should use programs designed to produce total communication competence.

We are in complete agreement then with Goodman when he concludes that one must start with the approach that reading is an aspect of language which in turn is an aspect of communication.

Because language is more than just knowledge of structure, and because the context of language usage is interrelated with the form of language, it is vitally important that we start with the communication competence of the child and build on this.

Cazden and Bartlett (1973) have attempted to point out why it is necessary to teach language skills as a component of communication competence and to delineate some of the important implications this holds for the teaching of language skills. They point out:

(1) That various codes, e.g. oral, written, restricted, elaborated and other classifications which have been proposed, draw on the same language system. The difference between them lies in the function they serve and the context in which they're used. Thus in a language program emphasis will be on changing patterns of frequencies of use of certain forms not in merely trialling new forms themselves. Curriculum design thus becomes a matter of creating tasks (pragmatic situations in Mt. Gravatt terms) which will elicit these forms - tasks which require the child to communicate, to define and express himself. In coding his ideas in particular contexts, the child will be extending (and indeed recombining his habitual syntactic patterns) to the learning of new ways of expressing himself. (It is hoped that in summarising Cazden and Bartlett's first point, their meaning intentions have not been distorted, and I have not read into their work our own meaning intentions). In any case it is rewarding to find a summary such as this (written by eminent language scholars from Harvard), which so closely matches the outcomes of our own research data to be described shortly.

(2) That a full language program requires language experiences which stimulate the functional use of language.

(3) That while the influence of momentary context on speech is acknowledged, "code" is distinguished from "speech variant". Code refers to a more pervasive or underlying regulation of speech. It is the child's code we wish to extend orally and transfer to the written and read language.

(4) That children need to discuss their ideas and plans with each other and their teacher. The result of this is planned individualised learning rather than everyone doing his thing.

(5) That language use in the total school environment, not just a particular period of the day should be the object of concern. Situations which can be created in school provide "concentrated encounters", give practice in new roles, and help some children overcome previous habits of coping with their world in non-verbal ways, or through restricted codes alone.

(6) That, in terms of content, the specific topics that communication is "about" should start with where the child is.

For teachers to implement such specifications in practice, they must know quite a deal about what they are to build on. They need to know what language units (chunks of language or language routines - we have many terms for this) children can use at any particular age, how they combine these units in expressive language and on what "rules" they operate when they associate language with meaning.

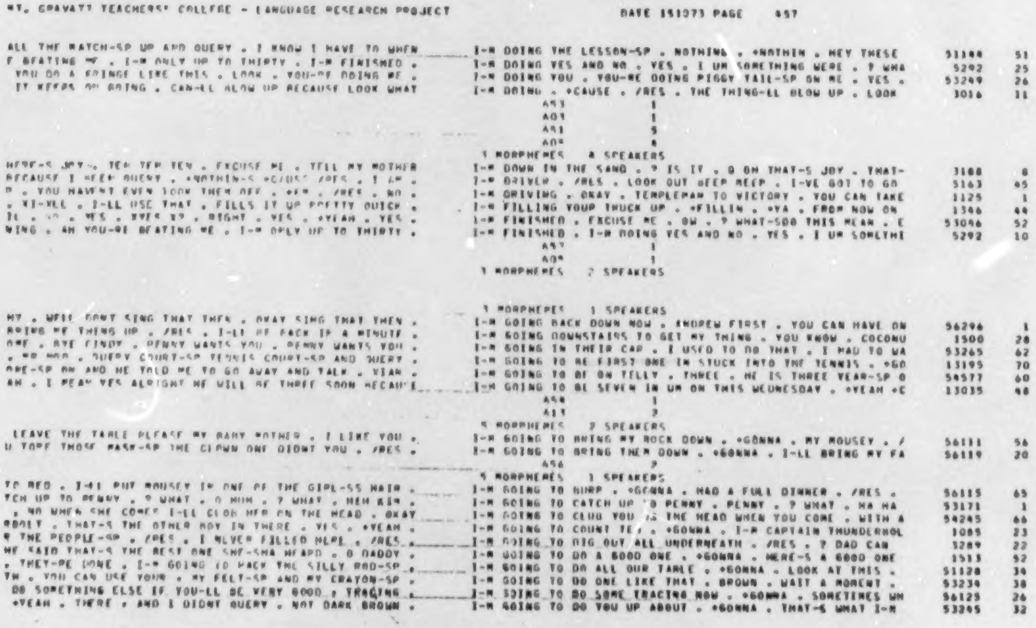
The Mount Gravatt Language Research Program attempts to supply some of this much needed information. One practical application of the research has become known through the publication by Addison-Wesley of the Mt. Gravatt Developmental Language-Reading Program; and the necessity to apply theory to practice in the building of that program has, in our opinion, been both a powerful motivator and, at the same time, a healthy brake on flights of research fancy.

We have concentrated, until fairly recently, on collecting a large oral and written language data base, on ordering that data and producing an apparent performance-type analysis which has competence implications. We have used these data and analyses, together with further semantic analysis as the language control for the language/reading program presently being published.

To build the data bank, we selected, for each of the ages 5½, 6½, 8½ and 10½ years, a random sample of all the children in Brisbane. We eliminated from each sample, by application of a language test, those children whose language ability appeared to be above or below the average for that age; and then collected a complete day's utterance from each of the remaining sample of children, the size of which ranged from 18 to 24. To do this, we had to pick up via radio microphone and a stereo cassette recorder, everything each child said and also the conversation of those to whom he was speaking, from when his day began until he went to bed at night. This involved, of course, following him from home to school, throughout the school day, back home, and then wherever he went until bedtime. All utterance, including the relevant speech of others was transcribed; and the child's utterance was coded for a computer concordance program which ordered it in such a way that all identical syntactic forms (we call these word sequences) over the entire age sample were brought together, one under the other, within the immediate verbal context of each occurrence. This is shown in Table 1.

TABLE 1

Sample Page of Computer Print-Out - 6½ Year Old Language



```
MT. GRAVATT TEACHERS' COLLEGE - LANGUAGE RESEARCH PROJECT                    DATE 151073 PAGE    457

ALL THE WATCH-SP UP AND QUERY . I KNOW I HAVE TO WHEN  I-M DOING THE LESSON-SP . NOTHING . +NOTHIN . HEY THESE      51188     51
F BEATING ME . I-M ONLY UP TO THIRTY . I-M FINISHED .  I-M DOING YES AND NO . YES . I UM SOMETHING WERE . ? WHA      5292     25
 YOU DO A FRINGE LIKE THIS . LOOK . YOU-RE DOING ME .  I-M DOING YOU . YOU-RE DOING PIGGY TAIL-SP ON ME . YES .     53249     26
 IT KEEPS ON GOING . CAN-LL BLOW UP BECAUSE LOOK WHAT  I-M DOING . +CAUSE . /RES . THE THING-LL BLOW UP . LOOK       3016     11
                                                               A53          1
                                                               A03          1
                                                               A51          5
                                                               A08          8
                                                        3 MORPHEMES   4 SPEAKERS
HERE-S JOY . TEN TEN TEN . EXCUSE ME . TELL MY MOTHER  I-M DOWN IN THE SAND . ? IS IT . O OH THAT-S JOY . THAT-      3188      8
BECAUSE I KEEP QUERY . +NOTHIN-S +C/US' /RES . I AM .  I-M DRIVER . /RES . LOOK OUT BEEP BEEP . I-VE GOT TO GO       5163     45
R . YOU HAVENT EVEN TOOK THEM OFF . +EM . /RES . NO .  I-M DRIVING . OKAY . TEMPLEMAN TO VICTORY . YOU CAN TAKE      1125      1
. XI-XLL . I-LL USE THAT . FILLS IT UP PRETTY QUICK .  I-M FILLING YOUR TRUCK UP . +FILLIN . +YA . FROM NOW ON       1346     44
IT . ?? . YES . EYES ?? . RIGHT . YES . +YEAH . YES .  I-M FINISHED . EXCUSE ME . OW . ? WHAT-SDO THIS MEAN . E     53096     52
WING . AH YOU-RE BEATING ME . I-M ONLY UP TO THIRTY .  I-M FINISHED . I-M DOING YES AND NO . YES . I UM SOMETHI      5292     10
                                                               A57          1
                                                               A08          1
                                                        3 MORPHEMES   2 SPEAKERS

                                                        3 MORPHEMES   1 SPEAKERS
MY . WELL DONT SING THAT THEN . OKAY SING THAT THEN .  I-M GOING BACK DOWN NOW . ANDREW FIRST . YOU CAN HAVE ON     56296      1
BRING ME THING UP . /RES . I-LL BE BACK IN A MINUTE .  I-M GOING DOWNSTAIRS TO GET MY THING . YOU KNOW . COCONU      1500     28
ONE . BYE CINDY . PENNY WANTS YOU . PENNY WANTS YOU .  I-M GOING IN THEIR CAR . I USED TO DO THAT . I HAD TO WA     53265     62
. MM MOO . QUERY COURT-SP TENNIS COURT-SP AND QUERY .  I-M GOING TO BE FIRST ONE IN STUCK INTO THE TENNIS . +GO     13195     70
ONE-SP ON AND HE TOLD ME TO GO AWAY AND TALK . YEAH .  I-M GOING TO BE ON TELLY . THREE . HE IS THREE YEAR-SP O     54577     60
AH . I MEAN YES ALRIGHT HE WILL BE THREE SOON BECAUSE  I-M GOING TO BE SEVEN IN UM ON THIS WEDNESDAY . +YEAH +C     13035     48
                                                               A58          1
                                                               A13          2
                                                        5 MORPHEMES   2 SPEAKERS
 LEAVE THE TABLE PLEASE MY BABY MOTHER . I LIKE YOU .  I-M GOING TO BRING MY ROCK DOWN . +GONNA . MY MOUSEY . /     56111     56
U TORE THOSE MASK-SP THE CLOWN ONE DIDNT YOU . /RES .  I-M GOING TO BRING THEM DOWN . +GONNA . I-LL BRING MY FA     56119     20
                                                               A56          2
                                                        5 MORPHEMES   1 SPEAKERS
TO BED . I-LL PUT MOUSEY IN ONE OF THE GIRL-SS HAIR .  I-M GOING TO BURP . +GONNA . HAD A FULL DINNER . /RES .      56115     65
TCH UP TO PENNY . ? WHAT . O MUM . ? WHAT . HEH KIM .  I-M GOING TO CATCH UP TO PENNY . PENNY . ? WHAT . HA HA      53171      1
. NO WHEN SHE COMES I-LL CLOB HER ON THE HEAD . OKAY   I-M GOING TO CLUB YOU ON THE HEAD WHEN YOU COME . WITH A     54245     66
BBOLT . THAT-S THE OTHER BOY IN THERE . YES . +YEAH .  I-M GOING TO COUNT TEN . +GONNA . I-M CAPTAIN THUNDERBOL      1085     23
R THE PEOPLE-SP . /RES . I NEVER FILLED HERE . /RES .  I-M GOING TO DIG OUT ALL UNDERNEATH . /RES . ? DAD CAN        3289     22
HE SAID THAT-S THE BEST ONE SHE-SHA HEARD . O DADDY .  I-M GOING TO DO A GOOD ONE . +GONNA . HERE-S A GOOD ONE       1513     52
. THEY-RE DONE . I-M GOING TO PACK THE SILLY BAD-SP .  I-M GOING TO DO ALL OUR TABLE . +GONNA . LOOK AT THIS .     51128     34
TH . YOU CAN USE YOUR . MY FELT-SP AND MY CRAYON-SP .  I-M GOING TO DO ONE LIKE THAT . BROWN . WAIT A MOMENT .     53234     38
 DO SOMETHING ELSE IF YOU-LL BE VERY GOOD . TRACING .  I-M GOING TO DO SOME TRACING NOW . +GONNA . SOMETIMES WM     56125     26
 +YEAH . THERE . AND I DIDNT QUERY . NOT DARK BROWN .  I-M GOING TO DO YOU UP ABOUT . +GONNA . THAT-S WHAT I-M      53245     32
```

An index following each ordering gave access to the full verbal context (including the utterance of other speakers), and to notation of the non-verbal context. In our 8½ and 10½ year old collection, situational information such as location, participants, actions and focus of the speaker, and other relevant data were recorded as a running commentary on one track of a stereo-recorder while the utterances were being recorded on the other track.

To add to the very large volume of data which has resulted, we have available, from previous research, similar corpora of 2½, 3½ and 4½ year old language. We have also collected written language, produced as spontaneously as we could arrange, for 8½ and 10½. We also have large corpora of language from Bristol, England, (collected by Gordon Wells), for Toronto, Canada, collected by Richard Handscombe of York University) and from black kindergarten children from the Watts area of Los Angeles (collected and provided by the South-West Regional Laboratories at Los Alamitos, California) and Australian Aboriginal children.

The concordance program we use also provides frequency counts for one-word, two-word and three-word sequences (syntactic ordering), the number and identity of speakers involved, and the percentage of corpus accounted for by each sequence.

## TABLE 2

## The Most Frequently Used Two Word Sequences in the Language of Australian, British and Canadian children

| AUSTRALIAN CHILDREN (BRISBANE) | | | | | | | | | | | | ENGLISH CHILDREN (BRISTOL) | | CANADIAN CHILDREN (TORONTO) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2½ Years | | 3½ Years | | 4½ Years | | 5½ Years | | 6½ Years | | 8½ Years | | 3½ Years | | 5 Years | |
| Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index |
| that's | 143 | I'm | 243 | I'm | 273 | I'm | 350 | I'm | 41[illegible] | I'm | 338 | I'm | 429 | and the | 357 |
| he's | 131 | I got | 125 | that's | 114 | I'll | 210 | I'll | 277 | I'll | 186 | I want | 370 | I'm | 135 |
| I want | 130 | that's | 111 | I'll | 87 | it's | 157 | it's | 168 | I don't | 175 | that's | 179 | I like | 118 |
| look at | 110 | 's a | 78 | it's | 80 | I don't | 144 | I don't | 149 | it's | 152 | it's | 141 | you know | 102 |
| in there | 76 | I don't | 71 | I got | 79 | I got | 126 | I've | 122 | I know | 146 | 's a | 127 | and I | 99 |
| there's | 54 | I'll | 67 | I want | 61 | I know | 101 | I got | 121 | I've | 136 | I got | 121 | I know | 56 |
| 's a | 47 | I want | 60 | you know | 38 | I can | 90 | that's | 98 | that's | 114 | I don't | 114 | and then | 54 |
| I got | 40 | I can | 52 | I can | 37 | I've | 76 | I want | 95 | I got | 108 | I've | 108 | it's | 48 |
| I'm | 39 | he's | 33 | 's a | 36 | you're | 74 | I know | 88 | you're | 86 | I'll | 99 | going to | 43 |
| that one | 39 | a little | 33 | I don't | 35 | got a | 72 | I have | 83 | I have | 68 | that one | 80 | I'll | 40 |
| it's | 31 | got a | 32 | in the | 27 | that's | 65 | you're | 83 | I was | 59 | I can't | 80 | and we | 39 |
| want to | 30 | it's | 29 | going to | 20 | I can't | 61 | I can | 73 | you know | 57 | 's that | 69 | I don't | 37 |
| in here | 26 | do know | 25 | I think | 18 | I want | 54 | you can | 57 | I can | 51 | I will | 53 | in the | 37 |
| he's (has) | 25 | going to | 25 | a big | 18 | I think | 41 | you know | 55 | I think | 51 | want to | 41 | and a | 32 |
| got a | 25 | in the | 22 | got | 16 | I didn't | 38 | I think | 49 | you got | 45 | I have | 40 | I was | 30 |
| look that | 23 | I think | 22 | one of | 16 | 's a | 36 | 's a | 40 | I can't | 45 | what's | 34 | I think | 29 |
| there you | 20 | do want | 18 | know what | 16 | you know | 31 | that one | 38 | I didn't | 45 | is that | 32 | and they | 26 |
| in the | 14 | I saw | 16 | I had | 16 | I was | 30 | I can't | 36 | you can | 44 | I do | 32 | to play | 22 |
| no I | 14 | look at | 15 | I know | 16 | I have | 29 | to go | 36 | 's a | 43 | I like | 31 | that's | 21 |
| what's | 13 | do you | 14 | I saw | 15 | I did | 29 | I go | 35 | I want | 37 | going to | 29 | we have | 21 |

## TABLE 3

## The Most Frequently Used Three Word Sequences in the Language of Australian, British and Canadian Children

| AUSTRALIAN CHILDREN (BRISBANE) | | | | | | | | | | | | ENGLISH CHILDREN (BRISTOL) | | CANADIAN CHILDREN (TORONTO) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2½ Years | | 3½ Years | | 4½ Years | | 5½ Years | | 6½ Years | | 8½ Years | | 3½ Years | | 5 Years | |
| Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index | Sequence | Index |
| I want to | 1986 | I'm going | 4999 | I'm going | 3707 | I'm going | 6126 | I'm going | 9531 | I'm going | 6050 | I'm going | 8758 | I'm going | 1037 |
| that's a | 1070 | I'm not | 807 | that's a | 793 | I'm not | 2960 | I'm not | 4171 | I don't know | 3519 | I want to | 6748 | you know what | 478 |
| he's going | 700 | that's a | 799 | that's what | 310 | I don't know | 1903 | I've got | 1669 | I'm not | 3096 | I don't want | 1264 | I like the | 432 |
| look at the | 480 | I don't know | 537 | I want to | 310 | it's a | 1146 | I don't know | 1499 | I've got | 1989 | I've got | 947 | and I like | 170 |
| he's (has) got | 295 | I got a | 434 | I'm making | 256 | it's not | 599 | I want to | 906 | it's a | 775 | that's a | 848 | I'm telling | 132 |
| he's a | 139 | I want to | 241 | you know what | 246 | I'll get | 531 | I'll have | 900 | I don't want | 693 | I want it | 658 | going to have | 118 |
| I got this | 107 | that's my | 175 | I'm not | 246 | I'll have | 509 | it's a | 744 | I'm a | 670 | what's that | 430 | I like to | 115 |
| there you are | 97 | I got to | 158 | that's the | 186 | I got a | 456 | I don't want | 660 | that's the | 571 | I want my | 376 | you know my | 99 |
| it's off | 84 | he's going | 156 | it's a | 116 | you're not | 454 | I'll do | 587 | that's a | 561 | I'm not | 363 | I don't know | 96 |
| want to get | 56 | I'm on | 114 | I don't know | 114 | I've got | 454 | it's not | 578 | it's not | 489 | I'm a | 339 | you know something | 93 |
| I'm going | 49 | it's a | 111 | I got a | 107 | I don't want | 375 | I'll be | 496 | I'll have | 455 | I'm coming | 290 | and I'll | 91 |
| it's a | 25 | I can see | 96 | 'm going to | 85 | I'm doing | 366 | I'm the | 473 | I'll do | 427 | I want the | 261 | and I'm | 91 |
| got one of | 16 | I don't want | 89 | I'll go | 79 | I want to | 333 | I'm doing | 448 | I'm doing | 382 | I'm just | 254 | to play with | 41 |
| in the box | 11 | that's the | 69 | I'll take | 63 | I'll go | 309 | I'll show | 416 | I got to | 372 | I'm going to | 235 | that's all | 36 |
| have a look | 9 | it's not | 49 | one of these | 52 | I'll put | 274 | I'll go | 382 | you're not | 352 | I don't know | 203 | we have a | 36 |
| I like this | 5 | he's a | 43 | that's all | 46 | I'm a | 260 | I'll get | 382 | I'm the | 318 | it's a | 190 | it's not | 35 |
| 's (has) got a | 4 | I'll have | 41 | I've got | 37 | I'm the | 252 | I'm getting | 381 | I'm up | 305 | it's not | 169 | going to get | 32 |
| 'm going to | .2 | I'll come | 31 | it's not | 32 | I got to | 207 | that's a | 378 | that's what | 294 | I'm having | 145 | I know what | 31 |
| 've got one | .01 | we're going | 30 | it's going | 32 | I got one | 205 | 'm going to | 370 | I'll get | 270 | I got a | 137 | and then we | 29 |
| | | 'm going to | 27 | I got gun | 32 | I'll show | 168 | I'm up | 348 | I don't think | 256 | I want a | 125 | going to play | 24 |

To produce language development and reading materials in which the language used was matched to oral language competence at each age level, we had to check every sequence of language against the corpus to determine whether we were using that unit in the way most children of that age normally used it. This has led us to undertake a systematic syntactic-semantic-pragmatic analysis of each language corpus and to compare corpora to detect developmental trends. There is little need to tell you that our task is somewhat frightening in its size and complexity.

The position, as we see it, is that the great variability of language across samples from different children has hitherto been an insurmountable obstacle to anything approaching a description of the general language competence of children at different ages of development or with different backgrounds. Linguists are still struggling to find a model of language functioning applicable to language analysis. We, and apparently others abroad, think that we have found a viable and valid way through the problem.

It is only fair to say that we have looked at the various viewpoints of linguists. We are in correspondence with a considerable number of those currently researching or theorising in language and language acquisition, and there is no doubt that we have been influenced by their views. There is only time at present, however, to point to differences rather than similarities between the conceptual framework we have adopted for language analysis and the perspective of others engaged in similar work.

Analysis of language, for many years, has largely centred on syntactic analysis with some lip service paid to other dimensions of language. The focus of analysis has moved recently from syntax to semantics. This is discernible only in very recent literature, but it is very evident in the work presently being carried out abroad. Latterly, also, the focus has broadened to include the external context of language, such as interpersonal purpose, pragmatics, and the non-verbal components of communicative events.

Many of the most recent researchers have confined their attention (for example the most recent studies of Nelson (1973), Braine (1971), Bloom (1971), and Bowerman (1973)) largely with the acquisition of words or word concepts by children of pre-school age. Others, such as Halliday (1972), Slobin (1973), and Wells (1975) have included situation, notably interpersonal purpose, within their focus as well as the syntactic and semantic dimensions. More importantly, the most important studies still seem to seek information about the language children use (usually to the extent of producing statistical analyses across categories) rather than finding out how children use language.

We are attempting to describe language and language development from a functional standpoint. It is our view, therefore, that the classification system we use must relate simultaneously to all dimensions of language. A tall order, I know, but, as I think Mr. Walker will convince you, it can be done. Let me try to tell you why it must be done. Like Halliday, Cazden and others, we believe that the primary function of language is to convey meaning - therefore, our analysis must have constant reference to the semantic dimension. But the meaning intended by a speaker and the form he adopts to convey that meaning are very much interrelated - semantic and syntactic analysis must be integrated. Moreover, a speaker shapes an utterance to convey his intended meaning only within the one unique situational context; and in doing so he usually relies on both language and non-language cues to meaning. It is our view, therefore, that context influences not only meaning that a speaker sets out to convey; but also how much of the cueing of meaning is encoded in actual verbal form, and, therefore, the actual form used to provide the verbal cues. That, briefly, is why we assert that syntax, semantics, and situational context (or pragmatics) are functionally integrated and must be so regarded when we analyse oral language. They are inextricably interrelated.

The final major principle we have adopted in functional language analysis, is that the mental processes related to language activity take place at two interrelated levels of functioning:

(a) the operative, decision-making level; and
(b) the integrative, more automatic level.

We are led to this belief by consideration of the limits of human cognitive processing and of the nature of human cognitive non-language behaviour. We all reduce larger and larger sequences of behaviour to an automatic level of functioning so that we can concentrate on higher-order operative decisions. That is how we master more and more complex tasks. This is an obvious but extremely important aspect of language functioning which seems to have been overlooked. No analysis which purports to reflect the realities of language functioning can afford to neglect it.

For those of you who wish to know more about this first stage of our research program I commend to you "The Language of Children: A Key to Literacy", Hart, Walker and Gray, Addison-Wesley, 1977. For those who wish to see how we have provided for practical outcomes from our research, the Teacher's Manuals of both Levels I and II of the Mt. Gravatt Developmental Language Reading Program are available.

Now that I have given you what might be called the specifications for a functional language analysis, I leave it to Mr. Walker to describe the product and show you how it is built into the required specifications.

## REFERENCES

| | | |
|---|---|---|
| Bloom, L. | 1971 | "Why Not Pivot Grammar?", *Journal of Speech and Hearing Disorders* 36, pp. 40-50. |
| Bowerman, M. | 1973 | *Early Syntactic Development*, Cambridge University Press. |
| Bowerman, M. | In press | "Semantic Factors in the Acquisition of Rules for Word Use and Sentence Construction", *Directions in Normal and Deficient Child Language*. University Park Press, Baltimore. |
| Braine, M. | 1971 | "The Acquisition of Language in Infant and Child" in C. Reed (ed.), *The Learning of Language*, Appleton-Century-Crofts, New York. |
| Cazden, C.B. | 1977 | Concentrated versus Contributed Encounters: Suggestions for Language Assessment in Early Childhood Education in *Language & Learning in Early Childhood*. Ed. by A. Davies. Heinemann Educational Books Ltd. London. pp. 40-59. |
| Cazden, C.B., and Bartlett, E. | 1973 | Review of Gahagan, D.M., and Gahagan, G.A. *Talk Reform: Explorations in Language for Infant School Children*, Sage Publications, Beverly Hills, Cal., 1970, in *Language in Society* 2, pp. 147-60. |
| Goodman, K.S. and Goodman, Y.M. | 1976 | *Learning to Read is Natural*. Paper presented at Conference on Theory and Practice of Beginning Reading Instruction, Pittsburgh. |
| Goodman, K.S., Smith, E. and Meredith, R. | 197[illegible] | *Language and Thinking in School*. 2nd Ed. Holt, Rinehart & Winston, New York. |
| Halliday, M.A.K. | 1972 | *Learning How to Mean*, Department of Linguistics & African Languages, University of Nairobi. |
| Nelson, K. | 1973 | *Structure and Strategy in Learning to Talk*, Monograph Society for Research in Child Development, Vol. 38, No. 149. |
| Slobin, D.I. | 1972 | "Cognitive Pre-requisities for the Development of Grammar" in *Studies of Child Language Development*. Ed. by Slobin, D.I. and Ferguson, C. Holt, Rinehart & Winston, New York. |
| Smith, F. | 1971 | *Understanding Reading - A Psycholinguistic Analysis of Reading and Learning to Read*. Holt, Rinehart & Winston, New York. |
| Wells, G. | 1974 | *Language Development in Pre-school Children*, School of Education, University of Bristol. (not published). |
| Wells, G. | 1975 | *Interpersonal Communication and the Development of Language*. Paper given at Third International Child Language Symposium, London. |

# AN ANALYSIS OF CHILD LANGUAGE IN USE - (PART II)

(Paper presented by R.F. Walker, Mount Gravatt C.A.E. at the
A.A.R.E. Conference, Canberra. November, 1977)

The central aim of the Mount Gravatt Language Research is to gain those insights into child language and language development which we consider will be of direct use to teachers and others who are concerned with children. What we consider important depends, of course, on our view of language and language functioning. For that reason, it is necessary to preface a description of our analytic framework with a brief outline of our theoretical focus.

Because we are seeking information which is directly applicable to the learning and teaching of language, our perspective has been a functional one. The recent great blossoming of theory building activity across the various fields of linguistics, psycholinguistics, and pragmatics has evidenced a remarkable general shift towards a functional perspective, but the various contributions often seem irreconcilable rather than supportive. What appears to be needed is an integrative conceptual framework within which ongoing theory-building and research will be comparable and complementary. Accordingly, we have attempted to place our research within a relatively comprehensive framework which integrates major relevant contributions from the several fields.

## BASIC CONCEPTS

The following concepts, taken from the fields of linguistics, psycholinguistics, sociolinguistics and pragmatics appear to be both relevant and important to our purposes. It seemed to us that, unless these could be refuted, they had to be included within any such integrative model:

(a) From psycholinguistics comes the notion that the mental activity associated with language functioning occurs at the integrative (semi-automatic) level as well as at the operative (executive) level. This is the view expressed in Dr. Hart's paper. It was effectively argued also, although on different grounds, by Miller (1957) as long as twenty years ago, and, to my knowledge, his argument has not been refuted.

(b) Language is, to quote Olson, (1975) first and foremost a means of communication for the "regulation of practical actions directed to the achievement of personally and socially valued ends". Probably the name most often associated with that viewpoint is Halliday. Again, there are a few, if any, dissentients among those who seek to study language in use. It follows that if the aim is to gain insights into how children use, and acquire the ability to use, language, the kind of language studied should be that associated with "directive" behaviour such as is used in normal everyday functioning.

(c) That the instrumental function of language within directive behaviour is to convey meaning has again been generally agreed upon. As Clark (1977) states, "without meaning there would be no real point in language". What meaning is and the identification of various kinds of meaning present problems, but the centrality of meaning in language as a vehicle of communication seems virtually undeniable.

(d) From the centrality of meaning in communicative language functioning, flows the assumption that to acquire language is to learn to map meaning onto language forms and the reverse. It is evident, however, that meaning and context, both linguistic and extra-linguistic, are closely interrelated. To investigate what a speaker means by what he says, it is necessary to have a knowledge of the context, both verbal and non-verbal, in which he is speaking. Moreover, such workers as Shugar (1973) have forcefully argued that the context in which language is spoken, and the context spoken about in language are both significant in determining the meaning encoded in that language. It seems essential therefore, to establish a view of the operative relationships between form, meaning and context before setting up a framework within which to analyse communicative language.

The above is a statement, in broad terms, of some of the major aspects of the theoretical perspective upon which we base our Phase 2 analysis of language corpora. The procedures I shall outline below should be interpreted against that perspective. Although, as I have observed, twenty years have passed since Miller enunciated the apparently undeniable conclusion that, in use, language must be processed in units larger than single words, no large-scale analysis of language other than our own has taken cognisance of the implications of his statement. For that reason, I feel that I should dwell a little longer on that particular concept.

## THE BASIC UNIT OF LANGUAGE

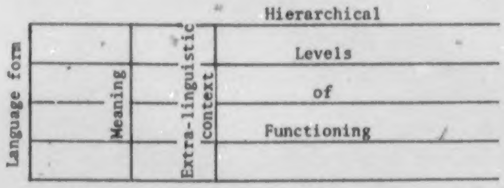
To accept this notion is to conclude that some, at least, of the units which language users process must be "chunks" larger than single words. This conclusion is a comfortable one because it is consistent with what we know of human behaviour outside language. No one would contend, for example, that a skilled tennis player makes a deliberate decision every time he moves a hand or foot; but rather that he calls up appropriate behavioural sequences which are woven into the purposeful behavioural stream. These sequences are much larger than the individual movements we might identify if we fully dissected a forehand drive, for example. A player in action does not and cannot attend to such miniscule elements. He restricts his mental activity at awareness level to the choosing of much larger behavioural sequences like the forehand drive and combining these in accordance with prevailing circumstances and his ongoing objectives. It seems evident that, in all areas of directive functioning, human beings attain higher levels of competence through reducing even larger behavioural sequences to what could be termed a semi-automatic level. Behaviour at this level involves mental activity at executive level only to the extent of monitoring the effectiveness of performance. Processing capacity is thus left available for decisions of a wider focus. When the limitations of human short-term memory are considered, it becomes evident, as Miller has pointed out, that fluent speakers do not make a conscious decision to add single word to single word. It follows that the basic units of language in use are not single words, but units larger than a word.

To deny that the word is the basic unit of language in use is to accept two obligations. The first is to identify alternate basic units for use in analysis of language; and the second is to investigate how the units identified are put together to encode meaning or meanings. We have the basis for the former in the Phase 1 Analysis, which Dr. Hart has described. The latter is the fundamental purpose of our Phase 2 analysis. At this stage, we consider that language corpora must be analysed in these two phases. Because children add to their repertoire of language units as they develop, this repertoire must be delineated for each corpus before its use can be examined across the various dimensions of language functioning. With Halliday, we agree that child language must be analysed against the system set up by the children concerned, not against an adult framework; so that, not only the repertoire of basic units, but also the rule system underlying their use may and probably will develop with maturation and experience.

Some of the segments of language, which children at a particular stage of development process as a single unit, coincide with what adults know as single words, and some are a string of words. Some of the latter, like "I dunno", or "Once upon a time", are used for only one purpose and tend to convey the same meaning, irrespective of context; but there are many word strings which occur frequently in children's language because they are functionally useful across a great variety of situations. Many of these can carry a variety of meanings from the very fact that they can be used in such a variety of contexts. Conversely, it seems reasonable to assume that those strings of words which are so functionally useful as to be used very often are the most likely to be used as one 'chunk' to the extent that, like the well-practiced behavioural sequences of the tennis player, they can be called up as a single unit and put into use by a single decision as to appropriateness. The full operational model is complex because, within the model, the uses for which these sequences are employed must be ordered hierarchically. The functional hierarchy is reflected on the syntactic dimension in that sequences and other "basic" language units combine to form "molecular" units which themselves combine at hierarchically higher levels of function; and so on. It can be assumed that units of all hierarchical levels are combined in accordance with operational rules of various kinds. Language development will involve progressive growth in this rule system, one which has reference, among other things, to language forms, meaning and extra-linguistic context and which interrelate those three dimensions of communicative language.[1]

---

[1] The operational relationship seen as existing between language form, meaning and extra-linguistic context is that diagrammed below.



At each hierarchical level, meaning is seen to arise from the mapping of language form (including intonation and stress in the case of oral language) onto an extra-linguistic context. In the production of language, form is seen to arise from the mapping of meaning (which relates to an extra-linguistic context) onto language form. As does the listener, the speaker also maps language form onto context; and so he monitors his own utterance against meanings intended.

## A HIERARCHY OF UNITS

To study these operational rules, it is necessary then to work out how language forms, at the various hierarchical levels, fit in with what is known about language and language functioning otherwise. This is a crucial issue because, if the syntactic rule system which speakers observe is not that of current grammars, (and it seems that it is not if only because the word, a unit on which they are built, is not functionally real) then a different system of rules, based on functionally real language units must be set up.

You will recall that the Phase 1 (concordance) analysis, described by Dr. Hart, identified frequently used word strings, and provided a convenient means of investigating the meaning behind any language unit - for any particular occurrence of it. A word string, as it occurs on the concordance page, is only the written form of that language unit; but it is useful to start with form, remembering that form, meaning and context must be interrelated.

"I'm going to" (usually pronounced by children as "I'm gonna") is one "sequence" which occurs frequently at all ages from 3½ to 10½ years. We can assume, then, that it is processed as a single unit, and that it is functionally very useful. For convenience, let us call it a "Sequence Unit". When we investigate the syntax of its use, we find that it is invariably combined with a variety of units of one particular type such as "swim", "play", "dig", "tell", and the like. Each of those units is a "contentive" in that it arises from and expresses some event conceived of as occurring within some particular extra-linguistic context.

We can also see that a "Sequence Unit", and one of those various "Content Units"[1] together make some sort of sense without further language context. We could name this syntactic unit - (a combination of a Sequence Unit and a Content Unit) a "Functional Unit", for want of a better term. It is useful to note that what we have called a Functional Unit is the smallest unit of language in the formulation of which it can be assumed that a decision for combination is involved; it would, therefore, be more appropriately termed a "Minimal Functional Unit" (M.F.U.).

"On the" is another frequently-occurring Sequence Unit. Also, it is used in combination with a variety of Content Units[2] - this time the type of content unit which refers to entities rather than to events. Some of the resulting Minimal Functional Units are "on the table", "on the grass", "on the wall" and "on the bus". Again the M.F.U. makes some kind of sense out of context. It appears that the addition of a Content Unit to a Sequence Unit narrows the range of possible extra-linguistic contexts onto which it could map. But a Minimal Functional Unit such as "I'm gonna sit" and one such as "on the grass" may be combined to form a larger unit of language such as "I'm gonna sit on the grass"; and there appears to be some difference in function between the two kinds of Functional Units. The first, "I'm gonna sit" carries a central thought; and the second, "on the grass" elaborates an element of that thought. We should then distinguish between two kinds of Functional Units, viz. Central Functional Units and Elaborative Functional Units at this molecular level, and perhaps also at other hierarchical levels of functioning.

We can also identify a language unit hierarchically above the Minimal Functional Unit and in that it is composed of Minimal Functional Units. We could term this unit an Utterance Sub-unit because it is the major constituent of full, extended utterances. For example, the utterance Sub-Unit, "I'm gonna sit on the grass" can be combined with other units of equal complexity to form an utterance such as "I'm gonna sit on the grass, but you're not allowed to because you've got a cold".

It seems, then, that there is a hierarchy of language units when we look at language form from a functional perspective, each unit being a combination of those on the hierarchical level next below it.

FIGURE 1

A HIERARCHY OF SYNTACTIC UNITS

| | |
|---|---|
| E | Interpersonal Level (Speaker 1 + Speaker 2 + ...) |
| D | Discourse Level (Utterance 1 + Utterance 2 + ...) |
| C | Utterance Level (I'm ..., but ..., because ...) |
| B | Utterance Sub-Unit Level (I'm going home / on the bus) |
| A | Minimal Functional Unit Level (I'm going + home) |
| (a) | Word (Morpheme) +... +... |
| (b) | Syllable + Syllable +... |
| (c) | Phoneme-grapheme +... |

---

1. "Pragmatic Terminal Unit" is probably a linguistically more accurate and informative term, but Content Unit is used here for simplicity.

2. Sequence units may also be combined with Minimal Functional Units such as in "I'm on the table", in which case the Functional Unit "on the table" can be seen to function as a Content Unit, suggesting a difference in functional level between the two Sequence Units "I'm" and "on the".

Figure 1 sets out some of the levels on which language can be analysed syntactically. These levels were established from an examination of our 5½, 6½ and 8½ year old corpora. It seems necessary to analyse at least up to Discourse Sequence level and down past what has been termed the Functional Unit Level to obtain a reasonably comprehensive description of the syntax of a child's language. It appears, too, that growth in language competence involves progress up and down the scale from the Minimal Functional Unit level as well as expansion on each individual level. Certainly, children come to break down "basic" language units into their constituent parts. It has been shown, for example, that children soon become aware of morphemes and other parts of words which carry meaning. The learning of the written form of the language hastens and eventually necessitates analysis down to the phoneme-grapheme level.

## SYNTACTIC-SEMANTIC ANALYSIS

Figure 1 could form the frame-work for syntactic analysis of language in use, but it had to be integrated with a semantic framework; and both had to be related to a model of language functioning if the Phase 2 analysis was to be in accordance with the view of language in use presented earlier. To return to the Sequence Unit, it is evident that "on the" carries different meanings in "on the grass", "on the wall", "on the bus", "on the second day", and in (a favourite one with children) "I'll hit you on the nose". It also seems likely that a child will use "in the" to convey some of those meanings before he uses it to convey others of those meanings. A use such as in "on the final occasion", for example, will certainly emerge relatively late. Besides noting growth in actual Sequence Units used at any stage of development, it is necessary then to look at growth in the meanings which children convey via the use of each individual Sequence Unit.

It is evident, also, that the particular meaning conveyed via the use of a particular Sequence Unit is determined to a considerable extent by the Content Unit with which it is combined in a Minimal Functional Unit. Consequently, it will be useful to note the nature of the Content Units which children combine with each Sequence Unit. It seems that some functional rule system must underlie the formulation of Minimal Functional Units. This rule system cannot be inferred without reference to meaning; and, as noted earlier, meaning cannot be determined without reference to context, both linguistic and non-linguistic.

An important interrelationship between the syntactic framework of analysis and meaning has already been noted. A Minimal Functional Unit such as "I'm gonna sit" was termed a Central Functional Unit because it conveyed the central thought (Central Concept) of that Utterance Sub-unit. If the Central Concept of an Utterance Sub-unit is adopted as a semantic unit, then we have a syntactic unit (Central Functional Unit) which corresponds with a semantic unit (Central Concept) in language functioning. From that starting point of coincidence, it is possible to construct a framework for semantic analysis of utterances which "mirrors" the syntactic framework. The analysis can then relate syntactic units, on all hierarchical levels, to corresponding semantic units. That the analysis be syntactic-semantic seems essential if it is to reflect the realities of language use - because it is evident that children do not operate on syntactic and semantic dimensions separately. They use some kind of integrative rule system.

FIGURE 2

AN INTEGRATIVE SYNTACTIC-SEMANTIC ANALYTIC FRAMEWORK
(SIMPLE UTTERANCE)

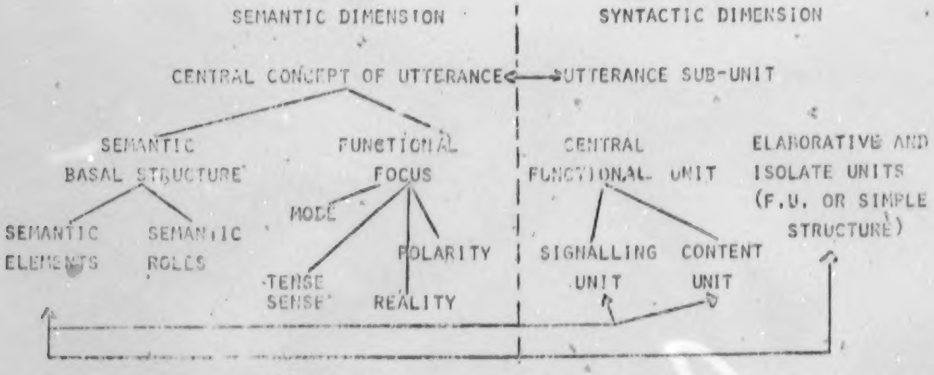


Figure 2 is a syntactic and a semantic analytic framework and some of the interrelationships between them. The terminology for semantic elements and roles was adapted from that developed in the field of generative semantics. This terminology is used to analyse the Central Concept of Utterance, referred to earlier as corresponding in language use to the syntactic unit "Minimal Central Functional Unit". There is considerable research support for the assumption that children develop, very early, the notions that there are Entities which have Attributes, that some Entities perform Actions, and that some Entities act on other Entities. Each Central Concept of Utterance, as defined in the analytic framework, can be seen as a semantic construct in which Entities, Events, and States are related to each other such that a State is attributed to an Entity, an Event is related to an Entity or Entities, and so on. The relationships conceived of by the speaker as existing between Entities, Events and States may be expressed by "Role" terminology ("Experiencer", "Agent", "Patient", "Benefactive", and the like).

The language units, or constituent parts thereof, which express the various semantic elements, or relationships between them, may then be identified. That is, the function of syntactic units can be described in semantic terminology. Elaborative Units, whether of Minimal Functional Unit or single-unit structure, can be seen to function as elaborating a particular semantic element of the Central Concept of Utterance. Such an integrated syntactic-semantic analysis may be applied at all hierarchical levels of the analytic model, starting at the Minimal Functional Unit level. When units of language are combined, another higher-order meaning is created, so that a hierarchy of meanings, corresponding to the hierarchy of forms can be distinguished. As Dewey has put it, "New meanings are bred from the copulating forms".

## THE CODING SCHEME

The Phase 2 coding scheme is designed to encode the syntactic and semantic elements and the relationships between them at a number of the hierarchical levels set out in Figure 1.

FIGURE 3

PHASE 2 CODING OF ONE UTTERANCE

(1st U.S.U.) (2nd U.S.U.)

*Level 1:* DEQSNAY QTHAT P1 EDAPCHY

*Level 2:* /1 IECH 4THERE 3S TEQX /2 4THAT TDPY 5COULD TABX EAZP EABU /

*Level 3:* /21 EABX 6FOR TOPI /

The information encoded can be interpreted in relation to Figures 1 and 2. Level 1 of the coding records the speech purpose of each utterance, the semantic analysis of each Utterance Sub-Unit, the semantic relationship between the Utterance Sub-units and the language unit, if any, used to express that relationship.

On the level 2 line, each syntactic unit, whether a word or a unit of larger size, is coded (in order of occurrence) for the semantic unit is expresses or elaborates, the category of entity, event, attribute, state or relationship it refers to, and some syntactic aspects of the unit. On Level 3, those language units from Level 2 which are of Functional Unit structure are analysed down to the basic language unit level. The coding of each of Level 2 and Level 3 records how each unit coded on those levels "fits into" the syntactic/semantic pattern of the level next above it.

## THE ANALYSES

A computer program has been developed to process the information encoded. The input program stores all the encoded information, and sub-programs are being progressively developed which obtain answers to individual questions as required. To the present, the 8½ year-old corpus has been coded and subjected to fairly considerable Phase 2 computer analysis. Coding of the 6½ year-old corpus has also been completed. Once the latter is analysed, we shall be able to see developmental differences, between 6½ and 8½ year-old language, in some detail over a fairly wide spectrum.

We are also carrying out Phase 2 coding of the language of matched pairs of close-contact Aboriginal and white Brisbane children at the time they entered the Mount Gravatt Developmental Language and Reading Program. This will be used for comparison with the language of the same children after some months in the program. Since the Phase 2 computer sub-programs present an analysis of the language of each child separately, as well as of a total sample, aspects of the longitudinal development of individual children will be revealed.

The Phase 2 analysis will no doubt need to be further refined and, possibly, expanded; but, once that has been done, it will be a comparatively simple matter to analyse our other corpora; the only evident problems being the training of additional coders (beyond the present group of four), maintaining a check on the consistency and accuracy of their work, and, of course, the funding of this activity. Outcomes of Phase 2 analyses are not yet available for dissemination; but the following tables illustrate some of the information which the analyses will yield.

(Tables and comment on the significance of some of the information will be presented)

## REFERENCES

Clark, H. & E. 1977 *Psychology and Language : An Introduction to Psycholinguists.* Harcourt Brace Javanovich, New York.

Miller, G.A. 1957 "The Magical Number Seven Plus or Minus Two : Some Limits on our Capacity for Processing Information". *Psychological Review.* 63, p81-97.

Olson, D.R. 1975 The Language of Experience - On Natural Language and Formal Education. *Address to Annual Meeting of British Psychological Society,* Nottingham.

Shugar, G. 1973 "An Ecological Method for the Study of Language Development". *Educational Review.* Warsaw Institute for Educational Research, 1, 2.

# GENERATIONAL DIFFERENCES IN ADOLESCENT INTELLECTUAL DEVELOPMENT [1]

Kenneth E. Sinclair
University of Sydney
1977

Rapidly changing social circumstances are often presumed to be having the effect of speeding up generational change in society. This speeding up, it is often argued is particularly noticeable with respect to the behaviour and development of adolescents. Evidence has been marshalled that indicates that the age of menarche in girls and that final weight and height in both boys and girls are reached at an earlier age relative to earlier generations (Meredith, 1963; Muuss, 1970; Tanner, 1962). This speeding up of physical development has been termed the "secular trend" and also appears to have its correlates in other areas of psychological functioning as well. Accelerated patterns of development across generations have been reported with respect to attitudes and interests (Jones, 1960; Connell et al., 1975) heterosexual activities (Broderick and Fowler, 1961; Harris, 1959; Havighurst, 1963; Kuhlen and Houlihan, 1965), intellectual functioning (Tuddenham, 1948) and even the legal definition of the attainment of maturity (e.g. voting age). The results of such investigations have served to stress the point that psychological development must be studied in the context of continuous social and technological change which not only helps determine the conditions in which youth behave and interact but also apparently influences the rate of physical and social growth.

Evidence relating to the secular trend or to general positive trends of accelerated development are not, however, completely without challenge. On the one hand, several overseas studies (McCammon, 1965; Poppleton and Brown, 1966; Himes and Maliva, 1975) have failed to find evidence for the secular trend. Similarly, in one recently published Australian study in which findings about age of menarche were compared with those of earlier studies, it was concluded that "...the Australian evidence is not strong enough yet to support an affirmative assertion of a secular trend for menarche" (Marder, Harvey & Russo, 1974). On the other hand, several studies have recently been published (Nesselroade & Baltes, 1974; McCandless, 1975) suggesting negative trends in development, for instance, declining performance of successive generations of students at school.

The results of research, therefore, are equivocal and even contradictory on the issue of generational change during adolescence. One problem in arriving at firm conclusions has been the lack of adequate data-gathering methods for the study of generational change (Sinclair, 1974). Most studies of developmental phenomena have used cross-sectional or longitudinal data gathering techniques which are not sensitive to the determination of cohort differences. Secular trend effects have most often been established by use of a time-lag design whereby scores of subjects at one point in time are compared to scores of a comparable group of subjects at a different point in time. For instance, high school students in 1975 are compared to those in 1950. A problem with such a design, however, is that developmental trends for particular cohort groups are not systematically established. In addition, it is also possible that time-of-testing differences observed may reflect the operation of a particular set of environmental conditions operating temporarily at one of the testing points.

The problems inherent in these three methods are largely overcome by the use of sequential designs advocated by Schaie (1965) and Baltes (1968) which are able to cope simultaneously with all three main components of development - age, cohort and time-of-measurement. Table 1 provides an example of Schaie's

Table 1

Schaie's General Model for Assessing
Time-Related Behaviour Change

| | | Time of Measurement | | |
|---|---|---|---|---|
| | | 1974 | 1975 | 1976 |
| Chronological Age | 12 | 1962* | 1963 | 1964 |
| | 13 | 1961 | 1962 | 1963 |
| | 14 | 1960 | 1961 | 1962 |

↓ Cross-sectional comparisons

↘ Longitudinal Comparisons

→ Time-lag Comparisons

*Year of Birth (cohort)

---

[1] This study is part of a wider study of Adolescent Development and Social Change. Grateful acknowledgement is given to support for the project received from the ARGC and to the staff and students of the schools participating in the project.

general developmental model. In the example, samples of subjects ranging in age (12, 13 and 14) are tested at three times of measurement (1974, 1975, & 1976). As can be seen from Table 1, within column comparisons conform to the conventional cross-sectional design, within row comparisons represent the time-lag design, and comparisons along the diagonal conform to the longitudinal design. Successive statistical analyses may be applied to different sections of the data collected in this way so as to separate out the influences of the different sources of developmental change and assess their magnitude.

Such designs have recently been used successfully in yielding evidence of a more convincing nature than previously about generational (cohort) differences and the effects of social change. The importance of generational change (cohort) effects on development may be most clearly seen in the results of Schaie's studies of differences in adult cognitive functioning (Schaie, Labouvie & Buech, 1973; Schaie & Strother, 1968). The investigators analysed various measures of intellectual ability taken at time intervals 7 years apart for subjects grouped at 7 year intervals from 25 years of age to 81 years of age. The subjects were members of a health contribution fund. Both cross-sequential and time-sequential analyses were made of the data. Very strong cohort effects were observed such that with age controlled more recently born cohorts achieved higher scores than cohorts born less recently. These cohort effects in fact were found to be more powerful than age related effects. The analysis indicated that the negative cross-sectional age gradients that had been recorded in studies (e.g. Jones & Conrad, 1933) with adult subjects were primarily attributable to cumulative and positive cohort effects favouring most recently born cohorts. When occasion-to-occasion differences within each cohort were considered there was a striking absence of any dramatic decline in ability with age suggesting stability of cognitive functioning throughout most of the adulthood range.

In a further study involving high school students (Baltes and Nesselroade, 1972; Nesselroade & Baltes, 1974) the investigation involved a 2 year sequential-longitudinal study of a number of measures of personality and ability. Longitudinal subjects representing year of birth cohorts for each year: 1954-1957 (ages 13-18) were tested in 1970, 1971 and 1972. The measures taken were the Cattell High School Personality Questionnaire, the Jackson Personality Research Form and the Thurstone Primary Mental Abilities Test. The analysis conformed to Schaie's cross-sequential (cohort and time).

For the personality measures it was concluded that historical, cultural change effects covering the period 1970-1972 were more prevalent than either cohort effects, which may have originated prior to 1970, or age differences. Specifically it was found that adolescents showed decreases in superego strength, social-emotional anxiety , and achievement during the 1970-1972 period independent of their age level. Similarly, through that period they also showed increases in independence.

Control analyses for the personality measures indicated that the repeated testing of subjects had had little effect and that subjects who had dropped out in the course of the study were not greatly different from those who remained in the study. These relatively non-significant results for retest and drop-out for the personality measures, however, were significant for the ability measures. When adjustments were made for retesting and drop-out effects it was concluded that the developmental gradients that remained reflected mostly age-related rather than historical-time-related variance. However, in some instances (e.g. number facility and verbal meaning), after the adjustments had been made, time-of-measurement effects remained significant and it appeared that 1972 adolescents tended to perform less well than their 1970 same age counterparts.

This finding of a negative cohort effect (more recently born adolescents not performing as well as less recently born adolescents with age controlled) with respect to certain ability measures is an interesting and challenging one so far as the impact of social change is concerned. It would not have emerged if a simple cross-sectional or longitudinal design had been used. The finding is consistent with other evidence from the U.S. showing a decline in the average Scholastic Aptitude Test score of high school seniors over the period 1962-1975 (McCandless, as reported by Zajonc, 1976). It is, on the other hand, contrary to the positive cohort effects established by Schaie in the studies already discussed (Schaie, Labouvie, & Buech, 1973; Schaie & Strother, 1968). The difference in results, however, may well be a function of the younger age group of high school students studied by Baltes. On the other hand, the significance of Baltes' study is obscured because of the strong test-retest effects recorded in relation to the ability measures, making his results equivocal. One means of controlling that difficulty would have been to draw independent samples within cohorts for the different measurement occasions. That is, instead of taking one sample from each cohort base and testing them at each time of measurement, independent samples could have been drawn from each cohort level for testing at each time of measurement. This in fact was the design used by Schaie & Strother (1968) and Schaie, Labouvie and Buech (1973).

The present study again seeks to explore the possibility of cohort differences in ability score with Australian subjects. Ability measures are taken at three measurement points spaced a year apart for a broad sample of Sydney high school students, this time using independent samples as suggested above. The study was part of a wider investigation of adolescent development using sequential data gathering methods.

## Method

The subjects were sampled from the population of students attending six Sydney metropolitan coeducational state high schools , and three feeder primary schools representative of three categories of living area as defined by the educational-ratio statistic* (two high schools and one feeder primary school for each living area). The data presented in this report are derived from the data gathering model presented in Figure 1.

| | | Time of Measurement | | |
|---|---|---|---|---|
| | | 1974 | 1975 | 1976 |
| | 6 | $C_3$ | $C_2$ | $C_1$ |
| School | 7 | $C_4$ | $C_3$ | $C_2$ |
| Years | 8 | $C_5$ | $C_4$ | $C_3$ |
| | 9 | $C_6$ | $C_5$ | $C_4$ |
| | 10 | $C_7$ | $C_6$ | $C_5$ |
| | 11 | $C_8$ | $C_7$ | $C_6$ |
| | 12 | $C_9$ | $C_8$ | $C_7$ |

$C_1$ - cohort commencing high school in 1977
$C_2$ - " " " " " 1976
$C_3$ - " " " " " 1975
$C_4$ - " " " " " 1974
$C_5$ - " " " " " 1973
$C_6$ - " " " " " 1972
$C_7$ - " " " " " 1971
$C_8$ - " " " " " 1970
$C_9$ - " " " " " 1969

Figure 1 : Data Gathering Model.

---

* Education ratio referred to the proportion of 15-19 year olds living in a particular local government area still engaged in full-time education. Local government areas in Sydney were ranked on the basis of this ratio with the top third categorised as education ratio (ER) I areas, the middle third as ER II areas and the lower third as ER III areas. These areas are assumed to approximate socioeconomic status divisions in Sydney in that, in an earlier factor analytic study, Jones (1968) had found that the education ratio statistic had one of the highest loadings on a factor interpreted as socioeconomic status.

---

On each testing occasion indicated (which was in April of each year) independent samples of students were randomly chosen from class and form lists provided by the schools. The ability measure used was the Otis Higher Test Form B. It was part of a battery of tests and questionnaires administered.

## Results

The data gathering design permitted two sequential analyses to be undertaken from the general developmental model of Schaie (1965). The time-sequential analysis involves levels of school year (6-12) x time (1974, 1975, 1976) x sex. The cross-sequential analysis involves cohort x time x sex. From the data gathering model it will be seen that cohorts 3-7 were tested on all three occasions thus giving 5 cohort levels in the analysis. In each case analysis of variance was used in establishing significant effects.

Table 2

Cross-sequential analysis of variance

| Source of Variation | Sum of Squares | df | Mean Square | F | Significance |
|---|---|---|---|---|---|
| Cohort | 55996.6 | 4 | 13999.2 | 14.9 | .001 |
| Time | 11591.4 | 2 | 5795.7 | 39.3 | .001 |
| Sex | 2948.3 | 1 | 2948.3 | 19.9 | .001 |
| Cohort x time | 2209.8 | 8 | 276.2 | 1.9 | .061 |
| Cohort x sex | 188.9 | 4 | 47.2 | .3 | .864 |
| Time x sex | 81.4 | 2 | 40.7 | .3 | .759 |
| Cohort x time x sex | 866.2 | 8 | 108.3 | .7 | .661 |
| Residual | 159509.7 | 1082 | 147.4 | | |

Table 3

Time-sequential analysis of variance

| Source of Variation | Sum of Squares | df | Mean Square | F | Significance |
|---|---|---|---|---|---|
| Sex | 1966.7 | 1 | 1966.7 | 14.5 | .001 |
| Year | 155796.6 | 6 | 25966.1 | 191.3 | .001 |
| Time | 1311.2 | 2 | 655.6 | 4.8 | .008 |
| Sex x year | 1211.7 | 6 | 201.9 | 1.5 | .179 |
| Sex x time | 315.3 | 2 | 157.6 | 1.2 | .313 |
| Year x time | 2040.4 | 12 | 170.0 | 1.3 | .242 |
| Sex x year x time | 1514.6 | 12 | 126.2 | .9 | .516 |
| Residual | 203308.6 | 1498 | 135.7 | | |

The results of the analyses of variance are presented in Tables 2 and 3. Table 2 contains the cross-sequential analysis and Table 3 the time-sequential analysis. From Table 2 it will be observed that the main effects of sex, cohort, and time of measurement were all significant. In general it was found that females were superior to males, that older cohorts were superior to younger cohorts, and that with each successive testing 1974-1976 performance improved. It must be recognised, however, that with cohorts 1 and 2, and 8, 9 excluded from this analysis (because they were not tested on all three occasions) both cohort and time-of-measurement differences are confounded with cross-sectional school year differences. Subjects in the older cohorts and tested at the most recent times of measurement are drawn from the later years of school. Thus the cohort differences and time-of-measurement differences may not be unequivocably attributed to generational or social change.

To interpret the result further, therefore, attention must be turned to the time-sequential analysis summarised in Table 3. Again all three main effects of sex, school year, and time-of-measurement are significant. In this analysis the school year effect is a very strong one with an F value of 191.3 and reflects student improvement with school year on the test. This is a much stronger effect than was recorded for the cohort effect in the cross-sequential analysis and, being closely related to age, suggests strong ontogenetic change. Again, however, because this effect is confounded with generational differences (students in the upper years of school originating from earlier school-year cohorts) it may not be unequivocably interpreted. In fact the significant time-of-measurement effect indicates that generational differences indeed are present. Furthermore, in this analysis, school year is controlled across all three times-of-measurement adding to the significance of the result. For students in years 6-12 performance on the test fell from 1974 to 1975 to 1976. This result reveals a negative cohort effect with older cohorts outperforming more recent cohorts with school year controlled. In sum, therefore, the analyses reveal significant effects for both positive ontogenetic and negative generational change although the former is the stronger effect of the two.

These effects may be further illustrated in Figures 2 and 3. Figure 2 presents the cross-sectional gradients for intelligence test performance and Figure 3 the within-cohort gradients. The cross-sectional gradients show generally that performance in 1974 over school years 6-12 was superior to that in 1975, which in turn was superior to that in 1976. These cross-sectional gradients are redrawn in Figure 3 to represent within-cohort gradients. For both males and females, for instance, it will be observed that the gradient for the students who began high school in 1974 (C4) is above that of the students who began high school in 1975 (C3) reflecting superior performance. Although there are some exceptions, this finding of older established cohorts outperforming more recently established cohorts reflects the general trend of the results.

## Discussion

In the present study sequential analyses were used to study the changing intellectual performance of students in years 6 to 12. Strong age related increments in intellectual performance were found both with respect to cross-sectional gradients that emerged and for gradients within particular school-year cohorts. Of greater interest and importance, however, negative cohort effects were also found such that performance showed a gradual decline across the school years involved from 1974 to 1976. This was also reflected in the cohort gradients with older cohorts being superior to more recent cohorts with school-year controlled. Thus, for instance, boys and girls who began high school in 1973 outperformed boys and girls who began high school in 1974 in the April testings of both 7th and 8th grade.

The present finding of negative cohort effects is consistent with findings in the U.S. of declining performance on the S.A.T. (McCandless, 1975). It is also consistent with the results of Nesselroade and Baltes (1974) who found that adolescents tested in 1972 on measures of number facility and verbal meaning performed less well than their 1970 same age counterparts.

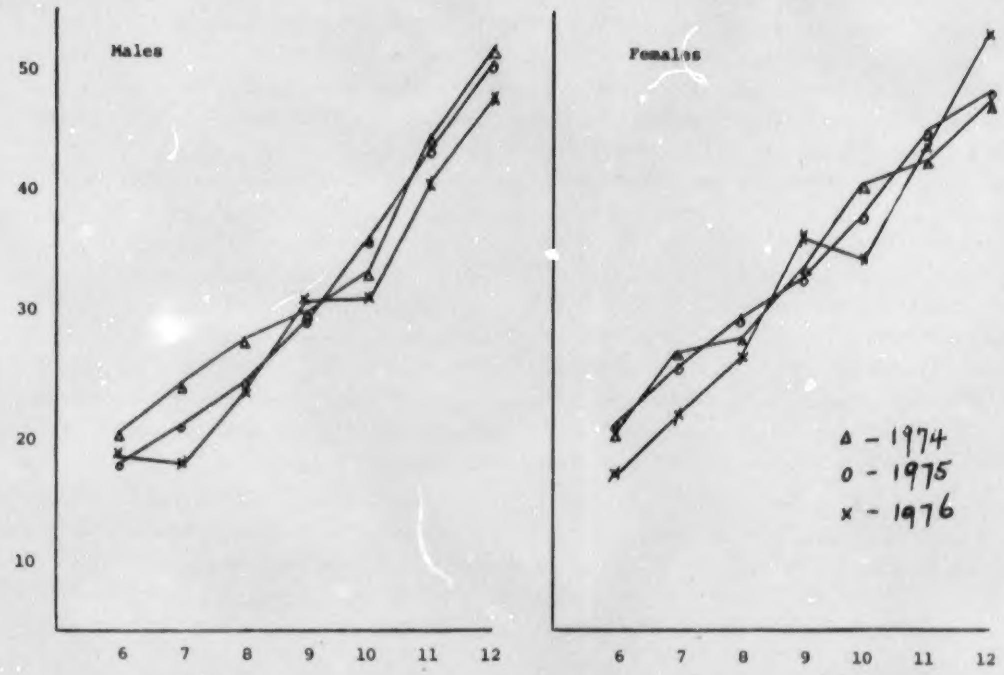


Figure 2 Cross-sectional school year gradients for intelligence test performance.

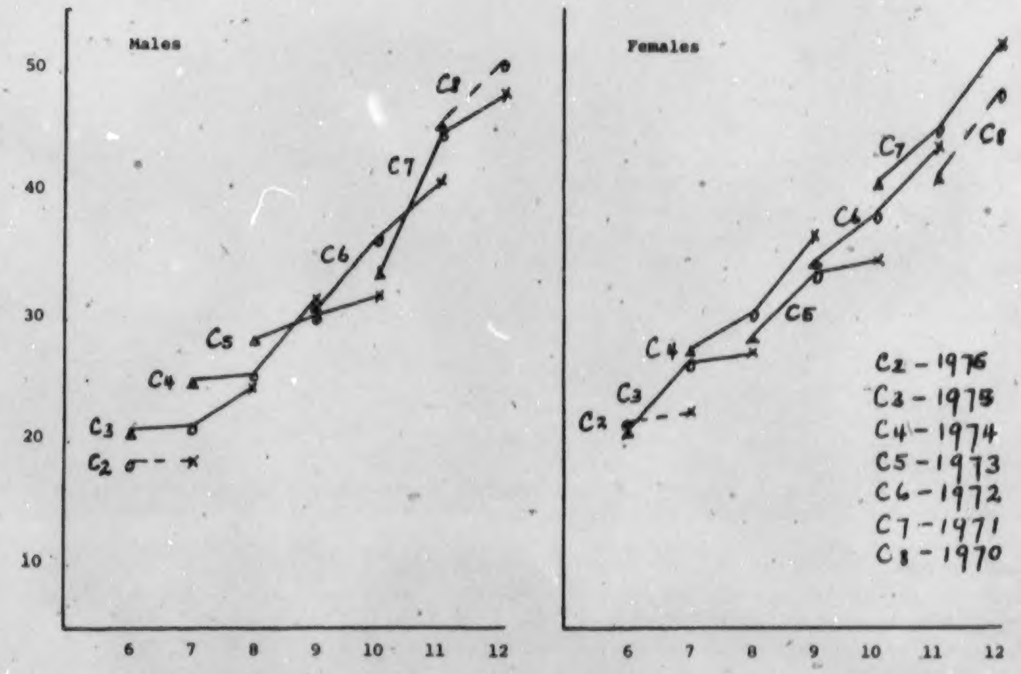


Figure 3 Within-cohort school year gradients for intelligence test performance.

The significant time-of-measurement effects, with school year controlled, found in the present study must largely be attributed to social change influences. The results themselves, however, are of little help in pin-pointing the specific social change influence involved. Some critics will undoubtedly attribute results to the effects of television, or to declining teacher commitment, or to the adoption of more progressive humanistic approaches to teaching, or even to the apparent decreasing classroom emphasis being given to tests and test-taking skills. As Zajonc (1976) has already argued, there is at present little evidence to support hypotheses such as these, although the need for research to test their validity is clearly evident. Zajonc (1976) finds more appeal in an alternative hypothesis. He presents evidence suggesting that negative cohort effects in cognitive performance may be attributed to changing patterns of family configuration relating to family size and birth order. The 12 year decline in S.A.T. performance (from 1963-75) parallels the increasing birth rate that occurred in 1946-57 (the birth years of the students tested in 1963-75). Zajonc argues on the basis of his confluence model that intellectual growth will be related to family size such that the larger the number of children the poorer will be the intellectual environment. The increase in family size in the period 1946-1957, therefore, is consistent with a gradual decline in intellectual performance of those birth cohorts when tested at age 18 in 1963-75. Furthermore, the increasing trend in family size after the second world war was reversed in 1963 and since then family size has gradually decreased. This leads him to predict that after 1981 (i.e. when the 1963 cohort is 18 years of age) S.A.T. scores should gradually rise once more.

Zajonc's hypothesis is an interesting one suggesting that perhaps similar trends and relationships could be established for the Sydney subjects of the present study. Further analyses in addition to this will also be undertaken with the present data. In one analysis year of birth will replace year of commencement at high school in forming cohort groups. In another, the same analyses used in the present study will be applied to scores on a test of formal operations which was administered at the same time as the Otis test.

..........

## References

Baltes, P.B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 1968, *11*, 145-171.

Baltes, P.B., & Nesselroade, J.R. Cultural changes and adolescent personality development: An application of longitudinal sequences. *Developmental Psychology*, 1972, *7*, 244-256.

Broderick, C.B., & Fowler, S.E. New patterns of relationships between the sexes among preadolescents. *Marriage and Family Living*, 1961, *23*, 27-30.

Connell, W.F., Stroobant, R.E., Sinclair, K.E., Connell, R.W., & Rogers, K.W. *12 to 20: Studies of city youth*. Sydney, Hicks Smith, 1975.

Harris, D.B., Sex differences in the life problems and interests of adolescents, 1935 to 1957. *Child Development*, 1959, *30*, 453-459.

Havighurst, R.J. Do junior high-school youth grow up too fast? *Bulletin of the National Association of Secondary School Principals*, 1963, *47*, 151-162.

Jones, F.L. Social area analysis: some theoretical and methodological comments illustrated with Australian data. *British Journal of Sociology*, 1968, *19*, 424-444.

Jones, H.E., & Conrad, H.S. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychology Monographs*, 1933, *13*, 223-298.

Jones, M.C. A comparison of the attitudes and interests of ninth grade students over two decades. *Journal of Educational Research*, 1960, *51*, 175-186.

Kuhlen, R.G., & Houlihan, N.B. Adolescent heterosexual interest in 1942 and 1963. *Child Development*, 1965, *36*, 1049-1052.

McCammon, R.W. Are boys and girls maturing physically at earlier age? *American Journal of Public Health*, 1965, 55,

McCandless, S.A. paper presented at the western regional meeting of the College Entrance Examination Board, San Francisco, 1975. Reported in Zajonc, R.B. Family Configuration and Intelligence. *Science*, 1976, *192*, 227-236.

Marder, K., Harvey, J., & Russo, P. The age of menarche in Sydney schoolgirls in 1973 with comment on the secular trend. *Australian Paediatric Journal*, 1974, *10*, 125-135.

Meredith, H.V. Change in the stature and body weight of North American boys during the last 80 years. In L.P. Lipsitt and C.C. Spiker (eds.), *Advances in child development and behaviour*, vol. 1, New York, Academic Press, 1963.

Muuss, R.E. Adolescent development and the secular trend. *Adolescence*, 1970, *5*, 267-284.

Nesselroade, J.R., & Baltes, P.B. Adolescent personality development and historical change: 1970-1972. *Monographs of the Society for Research in Child Development*, 1974, *39*, No. 1, (Serial no. 154).

Poppleton, P.K., & Brown, P.E. The secular trend in puberty: has stability been achieved? *British Journal of Educational Psychology*, 1966, *36*, 95-100.

Schaie, K.W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, *64*, 92-107.

Schaie, K.W., Labouvie, G.V., & Buech, B.U. Generational and cohort-specific differences in adult cognitive functioning: a fourteen-year study of independent samples. *Developmental Psychology*, 1973, *9*, 151-166.

Schaie, K.W., & Strother, C.R. A cross-sectional study of age changes in cognitive behaviour. *Psychological Bulletin*, 1968, *70*, 671-680.

Sinclair, K.E. Some methodological issues in the study of adolescent development. AARE Annual Conference, 1974, (mimeo).

Tanner, J.M. *Growth at Adolescence*. Oxford: Scientific Publications Ltd., 1962.

Tuddenham, R.O. Soldier Intelligence in World Wars I and II. *American Psychologist*, 1948, *3*, 54-56.

Zajonc, R.B. Family configuration & intelligence. *Science*, 1976, *192*, 227-236.

# ASSESSMENT OF SCHOOL ENVIRONMENTS

W.N. Bardsley
Western Australian Institute of Technology
on secondment to
Centre for the Study of Higher Education
University of Melbourne

AARE Annual Conference, Canberra, November, 1977

This paper is a draft only, submitted for discussion and comment prior to revision. It should not be quoted or cited in its present form without reference to the author.

## Abstract

Even to the most casual observer of secondary schools in Australia a variety of school environments is apparent. Differences in organisational structures at school level and in approaches to teaching and learning at classroom level are evident from school to school. Not infrequently the greatest contrasts occur between neighbouring schools. During the last decade schools based on flexible 'open plan designs' and schools incorporating 'open education ideas' have emerged whilst other schools have maintained a traditional approach to education. A pre-requisite for the study of the impact of different types of schools on students is a means of evaluating the school environment. This paper describes an attempt to construct and validate an instrument which may prove to be useful for the purpose.

Following on the work of Traub et al. and the independent effort of Walberg and Thomas a number of facets of the organisational and inter-personal aspects of secondary schools were identified and a questionnaire for teachers was devised. This 56-item Teacher Questionnaire was administered to a total of 154 teachers in 12 schools - two "open" and two "traditional" schools in each of three states (South Australia, Tasmania and Victoria). From an analysis of these data seven subscales were defined and the subscale scores were used to rank the schools along an open-traditional continuum. The results of this study confirm the dimensionality, reliability and validity of an instrument designed to assess school environments.

Any comparison of the effects of schooling leads naturally to consideration of the differences between (and within) the schools under investigation. One type of school may be said to influence positively the development of a particular attribute among students whilst another type of school could have a negative effect on otherwise similar students. It is often the case that arguments for or against a particular type of school are based on a subjective assessment of the school environment. However, it would seem to be nececsary to take account of certain key aspects of the school environment in order more adequately to understand the impact of the school on students. This paper seeks to identify those features of the environment likely to affect students and also likely to vary from school to school. Having identified several features the paper gives details of scales designed to measure each of them. In addition, an investigation of the reliability and validity of the instruments is undertaken. The relevance of the findings for both theory and practice are discussed.

## School Environments

Underlining the importance of the school environment for educational outcomes is the vast amount of money being poured into schools in Australia as a result of the Karmel Report (1973) with the aim, for example, of ameliorating adverse social conditions in disadvantaged areas (1973, Chapter 9), and fostering change at the school level by encouraging diversity and innovation in schooling (1973, Chapter 12). Examples of the kinds of projects funded through the Disadvantaged Schools Program are given in the Schools Commission Report for the triennium 1976-78 (1975, p. 62) and include the provision of specialist equipment, employment of personnel, construction of special facilities, running special classes, camps and excursions. The Special Projects Program (1975, p. 196) provides funds for grassroots innovations aimed at reducing traditional structures within the school, encouraging students to participate in making decisions about their own education and the running of their schools, meeting the needs of special groups, and other similar projects.

In addition to moves by the Federal Government to influence school environments a recent trend in Australian secondary education is towards alternative organizational structures at the school level and new approaches to teaching and learning at the classroom level. New high schools have been built incorporating "open plan designs" whilst some schools have introduced "open education ideas". One reason for this move towards open secondary education is that it is simply a further development of openness in primary education which developed as a result of the British influence and belief that the traditional school tends to dampen the child's natural enthusiasm for learning. Another reason is that secondary school students are increasingly insistent that their education be "relevant". Whatever the reason, in some secondary schools open education is a fact.

At the same time, and often in neighbouring suburbs, may be found the archetypal "traditional" secondary school holding steadfastly to traditional forms of organization and teaching methods. Likewise may be found other schools which appear to be uncertain of their status or direction. State secondary school students in Australia usually attend the high school that serves the locality in which they reside - there is virtually no choice of schools. In some respects the local high school may suit the student, in other respects it may not. Does this situation not constitute a challenge to researchers to investigate, amongst other things, the effects of traditional and open school environments?

But what is open education? It is a concept which seems to elude precise definition and almost every writer on the subject takes a different perspective. Nevertheless, there is some consensus and this would seem to be a good starting point for a discussion of open education.

Perhaps, in time, a theory of open education will evolve. At the moment there is no such theory and the best one can do is draw together some of the themes, attributes and practices that seem to characterize open education and contrast them with their counterparts in traditional education. By so doing, the concept of open education, as it is practised in some schools, may become more meaningful.

Generally, open educators are more concerned about the kinds of educational experiences students should have in school than about specifying units of learning, measuring outcomes and comparing performances (Bussis and Chittenden, 1972). It does not follow that open educators are unconcerned about what their students learn and how well they learn it. The teacher's role is to create and maintain the kind of supportive environment which best facilitates the learning environment (Lovegrove, 1974).

The extent to which open education is practised is reflected in school policy as well as in classroom situations. A number of these aspects of schools and classrooms have been identified by Traub et al. (1972) and are summarized below. Such characteristics are, of course, also seen in more traditional schools but to a lesser extent and to a lesser degree. For each attribute the properties of openness are briefly described.

## School Factors

1. Composition of classes. In an open school there are no "years" or "levels". Students are grouped without regard to age or past accomplishment. In such classes the students work individually or in small groups, free to talk and move about at will.
2. Assignment of students to teachers. This is a complex matter. However, schools which offer students a say in deciding which teachers they work with are relatively more open than those which do not.
3. Timetabling. In a school practising open education, there are no timetables, no bells and no compulsion for students to attend classes. Students begin an activity when they are ready for it and continue for as long as they like. The remaining time is independent study time during which the students are free to pursue their own academic interests.
4. Student control. The need for rules arises only when the behaviour of one or more students is adversely affecting other students. The responsibility for formulating rules to govern behaviour and for enforcing them is taken equally by the students and the teacher.

## Classroom Factors

1. Setting learning objectives. Since students develop at their own rate and in their own style it follows that objectives should be set for individuals not groups. In the open classroom students participate in setting both long-term and short-term objectives for themselves.
2. Materials and activities. In the open classroom the emphasis is on diversity of learning materials and educational activities. Something of interest should exist for each student. Some examples are teacher or commercially produced instructional kits, printed matter, audio-visual aids, educational games, apparatus and equipment.
3. Physical environment. The preference of open educators is for a flexible learning space. This should extend beyond the students' desks, beyond the classroom walls and into the community beyond the school.
4. Individualization of learning. Open education allows students the freedom to work at their own pace using methods of learning that they find satisfying and rewarding. The student is free to choose which learning materials he uses, the activities he engages in and whether he works on his own or in a group.
5. The role of the class teacher. In the open classroom the teacher rarely conducts a formal lesson. Instead the teacher-student interactions are informal yet concentrated, and occur individually or in small groups in response to ongoing activities. The teacher is a provider of suitable learning materials, a resource person and a counsellor. He structures learning situations. He encourages students to diagnose their own problems and formulate their own solutions.
6. Student evaluation. The main function of student evaluation in open education is to help the student decide what the student's objectives are and what the student does and does not know. This information can come from student self-reports and teacher observations. Little or no use is made of conventional tests in open education.

Traub et al. (1972) developed a 28-item Dimensions of Schooling questionnaire based on the above assumptions. They validated the scale by comparing responses from teachers in classrooms judged by experts to be models of open and traditional classrooms on the basis of the above characteristics. This instrument has a coefficient alpha estimate of reliability equal to 0.81 (1972, p. 81).

Another independent analysis of the concept of open education, which led to the construction and validation of an instrument to measure the properties of the classroom, has been carried out by Walberg and Thomas (1972, 1974). Based on a conceptual framework described by Bussis and Chittenden (1970), eight themes were selected by Walberg and Thomas (1972) as likely to distinguish between open and traditional schools. Items were collected and constructed to reflect each theme and subjected to a critical appraisal by prominent educationists. The revised items formed the basis of an Observation Rating Scale and a parallel Teacher Questionnaire each of 50 items. The scales were validated using both observational and teacher data from 21 United States traditional schools, 21 United States open schools and 20 open schools in Great Britain. Walberg and Thomas (1972, p. 206) claim to have "demonstrated a consistency among major analytic and descriptive writings on Open Education with respect to eight themes ... and to have found strong empirical evidence that the themes distinguish Open from Traditional teachers ..."

Summarizing from Walberg and Thomas (1974, pp. 7-9) the eight themes which are the basis of their instrument are presented below:

1. Instruction. Characterized by a very high degree of individual instruction and interaction in which an interdisciplinary approach is encouraged and the student's independence and exercise of real choice is fostered.
2. Provisioning. Many different learning materials are freely available within the classroom and the teacher encourages students to plan their own learning activities and share their learning experiences with other students.
3. Diagnosis. By observing and questioning the individual student as he goes about his activities the teacher is able to make judgements about the student's development that are more meaningful than information gained from standardized tests.
4. Humaneness. Within the classroom the products of the students' activities are displayed and used as learning materials.
5. Evaluation. Student evaluation is individualized and assessments are made on the basis of the teacher's observations with the aim of contributing to the student's development.
6. Seeking. The teacher seeks, uses and exchanges new materials, resources and ideas that will promote student development.
7. Self-Perception. The teacher prefers to deal with individual students as a source of stimulation rather than give formal lessons in set periods on specific subject matter.
8. Assumptions. The teacher believes in the innate curiosity of students, their eagerness to explore and their ability and right to make decisions about their own learning.

Whether or not the teacher's classroom behaviour is consistent with these eight themes in a basis for distinguishing between open and traditional classrooms. To assess this, Walberg and Thomas (1974) developed a 50-item observation rating scale. For each of the eight themes described above they constructed a number of items on which observers rated both open and traditional classrooms. The class teachers also rated their classrooms on the same items. The two methods were compared and both the observational and self-rating scales discriminated well and were highly reliable with internal consistencies of 0.95 or better (1974, p. 12).

## A School Environment Scale

That the school environment may be conceptualized in terms of a vast network of interrelated and interacting forces is not particularly helpful when attempting to delimit and define the various dimensions. Initially it may be useful to distinguish two broad, but by no means exhaustive, areas of the school, one at the organizational level and one at the inter-personal level.

Organizational Aspects. Fundamental to this aspect of the school are issues concerning school policies and organizational structure. Here such things as the placement of students into classes, drawing up of a master timetable, allocation of teachers to classes, subjects and rooms, student control and discipline, assessment procedures and after-school activities are considered. Schools vary widely in their approach to these matters. In some schools the principal may have all the say, in others the teachers may make decisions on such matters, and in others the student may decide for himself. Nevertheless, these are matters of deliberate policy which pervade the whole school.

Inter-personal Aspects. This aspect of the school concerns the interaction of the teachers and the students within the organizational structure of the school. Many factors enter into this but certain key issues can be identified. At the classroom-teacher-student level these include the teachers' approach to teaching and learning, whether or not students have autonomy with respect to their learning activities, the amount of stimulating materials and activities evident in the classroom, whether the teachers treat students as individuals or collectively, and the nature of the interactions between students in the classroom setting.

Of course, the school environment involves more than the structural and social aspects considered here. For example, such important matters as the standard of the school buildings and equipment, the curricula, regulations governing attendance, school hours, teachers' qualifications, staffing ratios, and external assessment all influence the environment of the school. However, in this study, they are regarded as falling within the domain of the State departments of education rather than the school. As the influence these aspects of the school have on the students is not likely to vary greatly they will not be considered further.

From the literature on school environments, (Traub *et al.*, 1972; Walberg and Thomas, 1974) items related to the organizational and inter-personal aspects were selected, modified where necessary and included in a questionnaire for teachers. One hundred and fifty four teachers in twelve schools (two "open" and two "traditional" metropolitan co-educational secondary schools in each of three states: South Australia, Tasmania and Victoria) completed the Teacher Questionnaire.

Data were collected in April, 1975. Depending on the size of the staff in the schools some twelve to fifteen teachers in each school were chosen to cover the whole range of subject areas taught in the school. To avoid the problem of teachers having different teaching styles with younger or older groups of students only teachers who most often taught students in their final year of compulsory schooling were invited to complete the Teacher Questionnaire. The overall response rate was 86 percent with from eight to fifteen teachers from each of the various schools returning the completed questionnaire.

These data were subjected to a series of principal component analyses and item analyses aimed at establishing the dimensionality of the instruments. The factor loadings are available elsewhere (Bardsley, 1976) and only the items defining the various subscales are given below. The item numbers relate to the original questionnaire which should be referred to for details of format. Items marked "R" are reverse scored.

School Organizational Structure Subscales. Seven items selected as indicators of the organization of classes or groups were found to constitute a uni-dimensional scale. These items provide data on the composition of classes, the variety of activities available at any one time, and the existence and requirements of a formal timetable. This subscale was called "Flexibility" as it indicates the degree to which organizational aspects of the school are adaptable to the needs and interests of the students, as opposed to the rigid timetable and class structure in some schools. The seven-item Flexibility subscale had an alpha coefficient of reliability equal to 0.77. The seven items are given below and the unweighted item scores, reversed where necessary, were summed to give the Flexibility subscale score.

Flexibility Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| 16 | Students are allowed to decide for themselves how much time they will spend on an activity | 0.51 |
| 29$^R$ | In any class students are about the same age | 0.42 |
| Amount of say students have | | |
| 35 | Deciding which class or group they will be in | 0.42 |
| 37 | Deciding what activities they will do in school time | 0.48 |
| Amount of each school day | | |
| 41 | Available for students to study whatever interests them | 0.62 |
| 42$^R$ | Students are required to spend on essential subjects | 0.40 |
| 45$^R$ | Organized into classes according to a set timetable | 0.64 |

Three items selected as indicators of the methods used to control students and establish discipline in the schools were found to be sufficiently intercorrelated to constitute a scale. The unweighted item scores for these three items were used as an index of the extent to which school rules are used to control students and secure discipline. Although the alpha coefficient of reliability of 0.54 is relatively weak, the three item scale was considered to be better than using a single item index.

School Rules Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| 7 | There are very few school rules | 0.45 |
| 11 | Students are mainly responsible for enforcing the school rules | 0.38 |
| 14$^R$ | Most teachers are very strict about school rules | 0.25 |

The Assessment subscale provides an index of the degree to which the assessment of students is formalized in the school. This three-item subscale has an alpha coefficient of reliability equal to 0.55.

Assessment Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| 21 | Student assessment is usually individually based and informal | 0.39 |
| Amount of say students have | | |
| 36 | Deciding how often they will be assessed | 0.51 |
| 40 | Deciding what method of assessment will be used | 0.48 |

Thus, three indices (Flexibility, School Rules and Assessment) of the school environment are available for comparing the twelve schools involved in this study in terms of specific aspects of their organizational structure. For each subscale a low score indicates that the teacher perceives the school to be open whilst a high score indicates that the teacher believes the school to be traditional in that respect.

Inter-personal Classroom Environment Subscales. With regard to inter-personal aspects of the school, twenty items were selected as indicators of the teachers' approach to teaching and learning, teachers' respect for the individuality of the students and methods and materials used by teachers. From a principal component analysis of this twenty-item Classroom Environment Scale four subscales were identified. The first subscale is defined by five items which, in the main, reflect a belief in the individuality of students. The five items are given below and when the unweighted item scores, reversed where necessary, are summed provide an index of "Individuality". This subscale has an alpha coefficient of reliability equal to 0.76.

Individuality Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| $26^{R}$ | Most teachers expect students in their class to work at the same rate | 0.53 |
| 32 | Many different learning materials are readily available to the students | 0.58 |
| 33 | Teachers assist students on the basis of the specific needs of the students | 0.52 |
| 34 | Students produce materials for use by other students | 0.43 |
| Amount of each school day | | |
| 46 | Teachers spend helping individual students or small groups | 0.59 |

The second subscale of the Classroom Environment Scale contains eight items which collectively suggest that informality is the key-note. Many different activities go on simultaneously in the classrooms and the students are free to choose which activity they will engage in, and with whom. Formal authority is minimized in favour of developing co-operation among the students. The eight-item raw score total was used as an index of "Informality" and the alpha coefficient of reliability of this subscale was 0.88.

Informality Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| 8 | Students are free to move around the room without permission of the teachers | 0.71 |
| $9^{R}$ | All students in the same room use the same texts, materials, equipment etc. at the same time | 0.51 |
| 10 | During class time students are encouraged to help one another with their schoolwork | 0.60 |
| 13 | Teachers often combine classes and work together as a team | 0.70 |
| 20 | In most rooms many different activities go on simultaneously | 0.78 |
| 28 | Within a class students form their own activities group | 0.63 |
| $30^{R}$ | The same method of finding out how much a student has learnt is used for all students in the same class | 0.70 |
| Amount of say students have | | |
| 39 | Arranging furniture and equipment in the classrooms or teaching areas | 0.55 |

The third subscale is defined by three items which indicate the extent to which the students are involved in the learning process. Whether or not they are encouraged to work things out for themselves and to evaluate their own efforts. The subscale defined by these items was called "Involvement". The alpha coefficient of reliability of this subscale was 0.70.

Involvement Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| 12 | Most teachers prefer to help students learn for themselves rather than present planned lessons | 0.63 |
| 19 | Students evaluate their own work and keep records | 0.44 |
| 25 | Most teachers let students work out for themselves their own methods of learning | 0.50 |

Four items which indicate the extent to which the teacher dominates and directs the activities of the class and, conversely, the students' lack of independence define the fourth subscale. It has an alpha coefficient of reliability equal to 0.75 and is called "Independence".

Independence Subscale

| Item Number | Item | Item - Total Correlation |
|---|---|---|
| In this school | | |
| $15^{R}$ | Most teachers use some form of class test to find out how much the students have learnt | 0.54 |
| $17^{R}$ | Materials are distributed and used under a teacher's direction | 0.48 |
| $18^{R}$ | Assignments are generally set for the class as a whole | 0.58 |
| Amount of each school day | | |
| $43^{R}$ | Teachers spend taking formal lessons with a class or group | 0.62 |

As with the three subscales used to evaluate aspects of the school organizational environment a high subscale score indicates that the teacher perceives the school to be traditional whilst a low score indicates that the teacher believes the school to be relatively open in that respect.

In the next section these four subscales (Individualization, Informality, Involvement and Independence) of the Classroom Environment Scale are used in conjunction with the three indices of aspects of the organization of the school previously described (Flexibility, School Rules and Assessment) to rank the twelve schools along an open-traditional continuum.

## Classifying the Schools

For each of the twelve schools the teachers scores on each of the seven subscales were averaged and the results of a series of one-way analyses of variance across the four schools in each State are presented in Table 1. For each subscale, school mean scores are presented in rank order from low (open approach of education) to high (traditional approach to education). In the table, figures underlined link together subsets of schools no pair of which have means that differ by more than the Tukey-HSD range value ($\alpha = 0.05$) for such subsets (Winer, 1962, pp. 77-89).

(Insert Table 1 about here)

When the means of the four South Australian schools (Nos. 1, 2, 3, 4) are compared it is clear that each subscale differentiates two open schools from two traditional schools. From an examination of the four Tasmanian schools means (Nos. 5, 6, 7, 8) it is evident that School 5 is significantly more open than the others on all but one scale (School Rules). School 6, although housed in conventional school buildings has introduced a number of innovations in an attempt to move away from the traditional mould. An indication of the partial success of these moves (but not in the area of Assessment) is shown in Table 1. Of the four Victorian schools included in this study, one (Number 11) was housed in an open space building whilst the others were all in older style conventional buildings.

From the results presented in Table 2 it may be seen that when all twelve schools are considered together, two groups, each of three schools, cluster together at opposite ends of the open-traditional continuum. Schools 1, 2 and 5 are obviously open schools in the sense used in this study whilst 3, 4 and 12 are clearly traditional . To confirm this finding and at the same time to differentiate between the remaining six schools. the following elementary ranking procedure was used. The twelve school means were ranked from low (1) to high (12) on each of the seven subscales and the rankings are presented in Table 3.

(Insert Tables 2 and 3 about here)

A measure of the extent of agreement between the rankingsof the schools on each of the seven subscales is given by Kendall's Coefficient of Concordance W (Siegel, 1956). For these data the high value of $W = 0.87$ ($p < 0.001$) provides evidence of interscale reliability and, as Siegel suggests "... their pooled ordering may serve as a 'standard', especially when there is no relevant external criterion for ordering the objects" (1956, p. 237). For the purpose of assigning a measure to each school along an open-traditional continuum Gaito (1959, p.119) claims that the average rank values obtained by this method approach interval level measurement. With these data it is possible to rank order the schools from the most open to the most traditional. It remains to consider the relevance of this procedure for practice.

## Implications for Theory and Practice

Turning now to the question of the validity of these results it is illuminating to review the procedure used to select the twelve schools. For the purposes of this study it was decided to select schools which differed considerably in their approach to teaching and learning as well as in their general organization. Following discussions with senior representatives of the State Departments of Education in South Australia, Tasmania and Victoria, four large co-educational metropolitan secondary schools were selected in each of these States. Essentially, the criterion for selection was that, in each State, two of the schools be typical "traditional" schools and two be schools using an "open" approach to education.

The two South Australian traditional schools (3, 4) selected by the Departmental officer were both long-established schools with reputations for academic excellence. Students attending these two schools were expected to wear the school uniform and nearly all of them did so. Lessons were conducted in conventional classrooms with desks in rows for students in established class groups. The school day was broken up into set periods for set subjects in set rooms with particular teachers according to a school timetable. Assignments were generally given to the class as a whole and assessment of students was by means of class tests and term examinations. In general, the teachers dominated the teaching-learning situation and the principal was supreme.

By way of contrast the two South Australian open schools (1, 2) selected for this study were designed as flexible open-space buildings. Although they had only been in operation for three years and were still in the evolutionary stage of development their primary concern was for the student as an individual. Students were expected to dress sensibly and were encouraged to wear the school uniform, but this was not a source of conflict between staff and students. The schools were well provided with equipment and resource materials and the teachers were also regarded as providing access to additional learning resources. Most teaching and learning took place in faculty areas with two or more teachers available to assist students as they worked individually or in groups using teacher-prepared printed cards. Less than half the school day was "scheduled time" according to a school timetable by means of which access to specialist areas was controlled. Student assessment was individually based, continuous and informal with no term examinations and few class tests. In general the student was encouraged to participate in making decisions about the learning process on the basis of his needs.

In Tasmania the traditional schools (7, 8) selected by the Departmental officer were similar in most respects to those in South Australia described above. One of the selected Tasmanian open schools (5) was built along open-space architectural lines while the other (6) was essentially a traditional building with traditional organization of classes but with an open approach to teaching and learning in some classrooms.

The two Victorian traditional schools (9, 12) were similar to the Tasmanian and South Australian traditional schools. As in Tasmania, one of the Victorian "open" schools (11) was an open space design while the other (10), a former "academic" high school, has since adopted a more informal style.

From the results in Table 3 it is evident that three "open" schools (1, 2, 5) and three "traditional" schools (3, 4, 12) define the end points of the continuum. Furthermore, schools 6, 10 and 11 form a group which may be considered relatively open or to some extent "innovative". Schools 7, 8 and 9 group together in what appears to be a relatively traditional position that might best be described as "conventional". It should be noted that the six schools originally selected as representative of "open" schools ranked from 1 to 6 using this method and the six "traditional" schools ranked from 7 to 12.

Although not a process study the four types of schools (traditional, conventional, innovative and open) described in this paper are not inconsistent with the stages in the development of an alternative school suggested by Deal (1975). He suggests that the initial drift away from the established pattern of education in the traditional school is seen in changing roles, expanded curricula and modified assessment procedures. This corresponds to the conventional schools in this study which, although still fairly traditional in most respects, are clearly much less formal in their dealings with students than is the case in traditional schools. The next stage in the evolution of an alternative school, according to Deal, is the development of a complex instructional program supported by a sophisticated organization. This stage is represented by the innovative schools in this study. They have a flexible organizational structure and involve students more in the learning process. In addition they promote independent learning by students. The final stage in the development of an alternative school is said by Deal to be reflected in a commitment to a completely individualized approach to learning in which the teachers and students have autonomy. In this study this corresponds to the open schools with their individualized approach to teaching and learning within a complex organizational framework designed to allow for individual needs.

In this paper the procedures presented for assessing school environments have been shown to possess both practical and theoretical relevance for educational research. Although no absolute criteria for classifying schools have been defined, the seven subscales described may provide a useful means of comparing the environments of different schools. Another application of these instruments might be to monitor the environment of a particular school over a period of time. Of more theoretical relevance, however, they might provide suitable measures of the school environment in investigations of the relationships between environment and student outcomes.

Table 1 School Environment Subscale Means and Significant Differences[a]

| Scale | | State | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | South Australia | | | | Tasmania | | | | Victoria | | | |
| | | Open | | | Trad. | Open | | | Trad. | Open | | | Trad. |
| Individualization | School | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 11 | 10 | 12 | 9 |
| | Mean | 8.2 | 10.8 | 14.9 | 16.5 | 9.7 | 13.0 | 14.8 | 15.7 | 13.8 | 16.1 | 16.3 | 17.0 |
| Informality | School | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 7 | 11 | 10 | 9 | 12 |
| | Mean | 14.1 | 19.3 | 33.2 | 33.2 | 14.4 | 26.7 | 28.0 | 28.3 | 23.0 | 27.2 | 27.7 | 32.8 |
| Involvement | School | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 7 | 10 | 11 | 9 | 12 |
| | Mean | 7.2 | 8.3 | 11.7 | 12.0 | 8.4 | 11.1 | 11.3 | 11.5 | 9.5 | 10.4 | 11.3 | 11.8 |
| Independence | School | 1 | 2 | 4 | 3 | 5 | 6 | 8 | 7 | 11 | 10 | 9 | 12 |
| | Mean | 9.4 | 11.3 | 17.2 | 17.2 | 11.5 | 15.3 | 15.6 | 16.6 | 14.5 | 14.8 | 16.5 | 17.4 |
| Flexibility | School | 1 | 2 | 3 | 4 | 5 | 7 | 6 | 8 | 11 | 10 | 9 | 12 |
| | Mean | 21.8 | 23.5 | 30.4 | 30.5 | 20.5 | 28.3 | 28.4 | 29.8 | 27.3 | 28.4 | 29.5 | 30.5 |
| School Rules | School | 1 | 2 | 4 | 3 | 5 | 6 | 8 | 7 | 10 | 11 | 9 | 12 |
| | Mean | 7.2 | 8.6 | 11.8 | 12.1 | 7.8 | 9.2 | 10.6 | 11.0 | 7.5 | 8.9 | 9.3 | 12.0 |
| Assessment | School | 1 | 2 | 4 | 3 | 5 | 7 | 8 | 6 | 11 | 10 | 9 | 12 |
| | Mean | 9.9 | 10.5 | 13.6 | 14.3 | 9.9 | 12.3 | 13.3 | 13.3 | 11.9 | 12.5 | 12.9 | 13.3 |

---

[a] For each scale each groups (States) school mean scores are presented in rank order from low (Open) to high (Traditional) together with the school identification number. The underlining links together subsets of schools no pair of which have means that differ by more than the Tukey-HSD range value for such subsets ($\alpha=0.05$). See Winer, 1962, pp. 77-89.

Table 2 School Environment Subscale Means and Significant Differences across all Schools[a]

| | | Open | | | | | | | | | | | Traditional |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individualization | School | 1 | 5 | 2 | 6 | 11 | 7 | 3 | 8 | 10 | 12 | 4 | 9 |
| | Mean | 8.2 | 9.7 | 10.8 | 13.0 | 13.8 | 14.8 | 14.9 | 15.7 | 16.1 | 16.3 | 16.5 | 17.0 |
| Informality | School | 1 | 5 | 2 | 11 | 6 | 10 | 9 | 8 | 7 | 12 | 3 | 4 |
| | Mean | 14.1 | 14.4 | 19.3 | 23.0 | 26.7 | 27.2 | 27.7 | 28.0 | 28.3 | 32.8 | 33.2 | 33.2 |
| Involvement | School | 1 | 2 | 5 | 10 | 11 | 6 | 8 | 9 | 7 | 3 | 12 | 4 |
| | Mean | 7.2 | 8.3 | 8.4 | 9.5 | 10.4 | 11.1 | 11.3 | 11.3 | 11.5 | 11.7 | 11.8 | 12.0 |
| Independence | School | 1 | 2 | 5 | 11 | 10 | 6 | 8 | 9 | 7 | 4 | 3 | 12 |
| | Mean | 9.4 | 11.3 | 11.5 | 14.5 | 14.8 | 15.3 | 15.6 | 16.5 | 16.6 | 17.2 | 17.2 | 17.4 |
| Flexibility | School | 5 | 1 | 2 | 11 | 7 | 10 | 6 | 9 | 8 | 3 | 4 | 12 |
| | Mean | 20.5 | 21.8 | 23.5 | 27.3 | 28.3 | 28.4 | 28.4 | 29.5 | 29.8 | 30.4 | 30.5 | 30.5 |
| School Rules | School | 1 | 10 | 5 | 2 | 11 | 6 | 9 | 8 | 7 | 4 | 12 | 3 |
| | Mean | 7.2 | 7.5 | 7.8 | 8.6 | 8.9 | 9.2 | 9.3 | 10.6 | 11.0 | 11.8 | 12.0 | 12.1 |
| Assessment | School | 1 | 5 | 2 | 11 | 7 | 10 | 9 | 8 | 12 | 6 | 4 | 3 |
| | Mean | 9.9 | 9.9 | 10.5 | 11.9 | 12.3 | 12.5 | 12.9 | 13.3 | 13.3 | 13.3 | 13.6 | 14.3 |

---

[a] For each scale school mean scores are presented in rank order from low (Open) to high (Traditional) together with the school identification number. The underlining links together subsets of schools no pair of which have means that differ by more than the Tukey-HSD range value for such subsets ($\alpha=0.05$). See Winer, 1962, pp. 77-89.

Table 3 Rank Order of School Environment Subscale Means across all Schools[a]

| | School | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Individualization | 1 | 3 | 7 | 11 | 2 | 4 | 6 | 8 | 12 | 9 | 5 | 10 |
| Informality | 1 | 3 | 11 | 12 | 2 | 5 | 9 | 8 | 7 | 6 | 4 | 10 |
| Involvement | 1 | 2 | 10 | 12 | 3 | 6 | 9 | 7 | 8 | 4 | 5 | 11 |
| Independence | 1 | 2 | 11 | 10 | 3 | 6 | 9 | 7 | 8 | 5 | 4 | 12 |
| Flexibility | 2 | 3 | 10 | 11 | 1 | 7 | 5 | 9 | 8 | 6 | 4 | 12 |
| School Rules | 1 | 4 | 12 | 10 | 3 | 6 | 9 | 8 | 7 | 2 | 5 | 11 |
| Assessment | 1 | 3 | 12 | 11 | 2 | 10 | 5 | 8 | 7 | 6 | 4 | 9 |
| Mean of Ranks | 1.1 | 2.9 | 10.4 | 11.0 | 2.3 | 6.3 | 7.4 | 7.9 | 8.1 | 5.4 | 4.4 | 10.7 |

[a] School environment subscale means given in Table 2 ranked from 1 (Open) to 12 (Traditional)

# References

Bardsley, W.N. "Student Alienation and Commitment to School: A Multivariate Analysis of the Effects of Home and School Environments". Unpublished Ph.D. thesis, The Australian National University, Canberra, 1976.

Bussis, Anne M. and Chittenden, E.A. *Analysis of an Approach to Open Education*. New Jersey: Educational Testing Service, 1970.

Bussis, Anne M. and Chittenden, E.A. "Toward Clarifying the Teacher's Role", E.B. Nyquist and G.R. Hawes (eds.) *Open Education: A Sourcebook for Parents and Teachers*. New York: Bantam Books, 1972.

Deal, T.E. "An Organizational Explanation of the Failure of Alternative Secondary Schools", *Educational Researcher*, 4:10-16, 1975.

Gaito, J. "Non-parametric Methods in Psychological Research", *Psychological Reports*, 5:115-125, 1959.

Lovegrove, T. "Recent Trends in Secondary Schools with Particular Reference to the South Australian Open-Space High Schools". Adelaide: South Australian Education Department, 1974.

Report of the Interim Committee for the Australian Schools Commission. *Schools in Australia*. Canberra: Australian Government Publishing Service, 1973.

Schools Commission. *Report for the Triennium 1976-1978*. Canberra: Australian Government Publishing Service, 1975.

Siegel, S. *Nonparametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill, 1956.

Traub, R.E., Weiss, J., Fisher, C.W. and Musella, A. "Closure on Openness: Describing and Quantifying Open Education", *Interchange*, 3:69-84, 1972.

Walberg, H.J. and Thomas, Susan C. "Open Education: An Operational Definition and Validation in Great Britain and the United States", *American Education Research Journal*, 9:197-207, 1972.

Walberg, H.J. and Thomas, Susan C. "Defining Open Education", *Journal of Research and Development in Education*, 8:4-13, 1974.

Winer, B.J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1962.

## SUNDAY 13 NOVEMBER

## 9.00 - 9.45am

| PAPERS | Rooms |
|---|---|
| J. Lawry (Gippsland I.A.E.)<br>Colonial Curriculum Development and Evaluation: Queensland 1859-1904 | Gateway Room |
| WORKSHOP ON CLASSROOM STUDIES | Seminar Room A |
| C. Marsh, K. Treharne (Murdoch University)<br>Responsive Evaluation of a Primary Schools Simulation Project (ERDC) | Ladies Drawing Room |
| R.W. Marsh (Victoria University of Wellington)<br>The Probable Limits of Early Intervention Programmes for the Substantially Intellectually Retarded | Seminar Room B |
| G. Trebilco (Rusden S.C.V.),<br>E.P. Atkinson (Monash University)<br>Trans 6/7: The Transition of Students from Primary to Secondary School | Meeting Room |

AARE Annual Conference 1977

# Colonial Curriculum Development Evaluation - Queensland 1859-1904

Dr J.R. Lawry

School of Education
Gippsland Institute of Advanced Education
Churchill Victoria 3842

# Colonial Curriculum Development and Evaluation - Queensland 1859-1904

J.R. Lawry

## Introduction

In 1860 when the Queensland Board of General Education adopted in its integrity the system of national education then in operation in New South Wales[1] the problems associated with curriculum development and evaluation were simultaneously resolved and avoided. The Education Act of 1860 provided no additional direction as to the content and organization of education. It must be concluded that, in the view of the government and the Board, the existing system satisfied the educational requirements of the new colony.[2] Few changes took place in the basic elementary curriculum until about 1890 when overseas practices and innovations were regarded as appropriate. By 1890 a narrow three R's dominated curriculum, designed to produce literate citizens whose further education, if any, was their own responsibility, was under pressure from several directions. There was increasing concern about curriculum revision, technical education, child development through the kindergarten method, and the national importance of education. The main source of information and criticism were several English Royal Commission Reports, including the Devonshire Report, 1872-75, the Samuelson Report, 1882-84, and the Cross Report, 1885.

After the worst effects of depression retrenchment and forced economies of the 1890s, the main critics were teachers' associations, which were gaining strength, and members of parliament representing the Labour interests, who were first elected in Queensland in 1893. Teachers took up educational issues in their own journals and the daily press. By the end of the 1890s disquiet about the future capacity of elementary education to meet demands for trained ability contributed to another review of the curriculum.

Two related issues, the failure to develop a satisfactory link between elementary or primary education and secondary education, and the inadequacies of technical education, attracted most attention in Australia until innovations in content and method, postponed by the depression, were incorporated in new primary curricula in England and Australia from 1900 onwards. The curriculum revision movement was assisted by the passing of the English Education Act of 1902 and the subsequent Elementary Code and Regulations for Secondary Schools of 1904.

In general the Queensland education system developed under the influence of the practical administrative problems of providing elementary education for an ever-increasing school population. There was popular support for the narrow elementary curriculum and teaching methods based on the theory of mental discipline advocated by the administrators, who were reluctant to take any initiative in the educational and administrative systems. The state's responsibility for education continued from 1859 to 1904 without any substantial changes, because a narrow elementary education was considered appropriate for the education, provided free after 1870, available to all children.

## Curriculum Development and Evaluation under the Board of General Education 1860-1875

Greater detail on this period is provided in Spaull, A.D. (ed) Australian Teachers, Macmillan, 1977, chapter 4, "Teachers and Task under Queensland Board of General Education, 1860-75."

The main features of the primary curriculum in the period 1860-75 may be stated simply.

1. The curriculum, published in the Queensland Government Gazette, 4 February 1861, provided for reading, writing, arithmetic, grammar, and geography - the ordinary branches of an English education - plus general religious instruction.
2. In some larger schools additional subjects, including Latin, French, German, Euclid and Algebra, were taught until removed finally by the Education Act of 1875.
3. Few changes were made to the official curriculum. Drawing was not prohibited, provided it did not interfere with the basic curriculum in schools seeking its introduction.[3] Playground apparatus, games and quasi-military drill were encouraged along with a graduated course in needlework[4] by some of the school inspectors. Basically the Board did not move to restrict the scope of the curriculum until this became expedient because of political pressure.
4. Political opposition to subjects such as Euclid, practical mensuration, rudimentary geology, botany, chemistry, mineralogy and mechanics[5] found support from some officers of the Board who favoured the narrow English education without extra subjects.[6]
5. Under the Education Act of 1875 the narrow view of the curriculum prevailed with the exclusion of Latin, mathema[illegible] and drawing from the primary curriculum. The General Religious lesson was also removed, and singing was included.

The Education Act of 1875 was the result of political expediency directed towards settling the education question by establishing a responsibile department to administer a narrow elementary and secular system of education. In the period 1860-1875 curriculum development and evaluation responded to some minor professional initiatives in providing or removing subjects not specified in the 1861 Regulations. The curriculum in force from 1876 was politically approved as being consistent with the responsibilities of the state in providing free elementary education.

## Curriculum Development and Evaluation under the Department of Education 1876-1904

The subjects of instruction provided for in Section 22 of the Education Act of 1875 and in the Regulations were reading, writing, arithmetic, English grammar, geography, history, elementary mechanics, object lessons, drill and gymnastics, vocal music, and, in the case of girls, sewing and needlework. Reading, writing, arithmetic, object lessons, drill and gymnastics, and vocal music were required at all levels, but needlework and geography were introduced in the Second Class, grammar in the Third Class, and history and mechanics in the Fifth Class.

The approach to school work in 1876-the development of mental culture-remained a dominant feature of the curriculum, being incorporated in the new schedule of 1904. The development of mental culture was held in such high regard that the adverse reports of inspectors on the results obtained in the formal subjects were ignored, and suggestions for practical subjects were put aside as inappropriate in the primary school. Added to formal school work was the inculcation of "the virtues of obedience, self-restraint, self-denial" through organised playground games. Work and play combined to assist the development of "habits of discipline and application to duty", which was one of the chief designs of school work.[7] The first Minister for Education, S.A. Griffith, supported a curriculum limited to "a sound elementary school education"[8] although some members of parliament had advocated vocational trade studies and agricultural education.[9]

Some changes were made as the result of experience in the schools. Unsatisfactory results in the upper classes in mathematics, grammar, geography, history, history, physics and drill contributed to some curriculum changes in 1881 mainly in eliminating the work in mechanics for girls in the Fourth and Fifth Classes and substituting laws of health and domestic economy, and in modifying the course in Third to Fifth Class needlework. No attention was given to the suggestions that, instead of preparing pupils for clerical positions, technical and agricultural education should be fostered, and cookery instruction given to girls to fit them as "domestic servants of wives of the working people."[10] The Department declined to introduce cumpulsory military drill, although prominence was given to drill in school work and few objections were anticipated to the formation of volunteer cadet corps in State schools.[11]

Criticisms of the curriculum were rare and usually ignored even when made by district inspectors and interested politicians. However senior administrators exerted undue influence. Music teaching was conditioned by the unwillingness of the Department to encourage teachers to devote much time to it[12] and by the dogged insistence over twenty-two years by Ewart, the General Inspector, that the old notation system be followed. In his opinion the sol fa system had not been "accepted and accredited by the musical world"[13] although it was advocated by teachers.

By 1885 the Department was beginning to meet some of the criticisms about the lack of practical subjects. Gymnastics instruction was introduced and plans were made to issue books on the Principles of Agriculture to schools for 1886 so that agriculture could be treated as part of the object lesson.[14] These were only slight concessions alongside the perpetuation of the disciplinary restraints of the intellectual subjects, and of drill which was prized as the means for securing habits of prompt obedience,[15] and later as a contribution to national defence. New Regulations in 1885 raised criticisms from inspectors and politicians. The claims of drawing put by inspectors were ignored.[16] Some members of parliament wanted more attention to basic subjects, others advocated more history for civic competence; Bible reading to counter vice, ignorance and larrikinism; and agricultural education.[17] Griffith delivered a justification of the mental disciplinary theory in response to its critics, although he later admitted that little resulted from the annual debate on the curriculum.[18]

The years from 1888 until 1904 saw a gradual accommodation to changes forced on the Department, during a period of economic depression and recovery, by the need to up-date the curriculum in accordance with practices in other colonies and by demands for higher primary, technical and secondary education.

In 1888 the kindergarten system of training was mentioned as a possible development along with drawing and evening classes for technical instruction.[19] Under Secretary Anderson and Ewart acknowledged for the first time in 1889 that changes had occurred in other colonies, and then they proposed or made changes. Temperance lessons were introduced as part of object lessons, drawing was proposed, and drill seemed likely to lead on to the formation of Cadet Corps in State schools. The first steps were taken with the Kindergarten method in the infant schools by organising a course of lessons by Mrs Goulden, who had successfully introduced the method in Victoria. Manual training in primary schools was kept under review as it had been established in London and Sydney. Vague hopes were expressed that technical schools would provide an attractive alternative to purely literary studies,[20] but the breakthrough came in subjects with commercial rather than technical value. In 1890 and 1891 several teachers were given permission to conduct private classes in book-keeping and shorthand,[21] but technical education continued to be ignored by the Department.

Administrative and curriculum changes were considered at departmental conferences in March and December 1889, but the revised Regulations did not come into force until 1 January 1892.[22] Extensive discussion of curriculum changes in the years 1889 to 1891 produced results confined to prescribing new readers, renaming the existing classes to introduce a Sixth Class, introducing temperance and first aid lessons under object lessons, raising the general standard of work required, and the up-grading of drill. It was suggested that the changes to the general course of instruction were designed to give more scope "to intelligent teaching in such subjects as reading, arithmetic, and grammar, from which accrue the highest practical advantage to the future everyday life of the pupil, as well as the intellectual gymnastics necessary for bracing the mind to logical and continuous thought."[23] The main change in the arithmetic programme added elementary geometry for boys, while girls devoted more time to needlework. The geography course was lightened to avoid the cramming of names, and history was restricted to the subject matter in the Royal Readers.

After a tour to the southern colonies in 1889 Ewart had suggested adding the study of composition to grammar teaching, and "mathematics and experimental science, as being at once more useful and affording scope for mental culture and development."[24] This line of development in the mental disciplinary theory was the only significant change accepted by Anderson and Ewart, but it merely imitated current practice wherever the theory was applied to primary education. The work in drill was up-dated by replacing the Schoolmaster's Drill Assistant with the revised Infantry Drill of 1889, used by the Defence Force.[25] The final feature of the new courses was the provision of the half-yearly programme for each subject which gave "definiteness and uniformity to the teaching and examination of the work."[26] Adjustments to the curriculum were made slowly, and, in the case of the Kindergarten method, postponed until a competent instructress was obtained.[27] Ewart showed little enthusiasm for the Kindergarten method as it had become "an educational fad or craze."[28] The Department believed that the changes to regulations and courses still left untouched the ordinary elements of an English education "necessary for boys and girls ... in any rank of life, to fit them for taking an intelligent part in the duties of citizenship." This programme, according to Ewart who ignored criticisms in parliament, met the expressed wishes of the population, and no alteration could be contemplated until there was a general expressed wish for a change. "In the meantime ... so far as is known the Department's work" was in accordance with the people's desire.[29] Ewart resisted such efforts to review the curriculum as appeared in stray articles and letters in the newspapers by seeking refuge in the argument that the Department was "administering the Education Act in a generous spirit and with even-handed justice all over the country, as far as circumstances permit and as far as money will go."[30]

The curriculum in force from the beginning of 1892 remained in operation with only minor changes until the new schedule of 1904 was introduced in 1905. The education offered was "largely utilitarian rather than ornamental ... a programme not for one class but for all classes of the community,"[31] in which most of the time was devoted to the basic subjects - reading, writing, arithmetic - and only minor attention was given to other subjects, except grammar. By 'utilitarian' the Department did not mean vocational or technical education because the primary education was clearly regarded as the minimal foundation upon which life and further education were built.

After several years of talk about the virtues of drawing, the Minister decided to include drawing in the curriculum under the subjects of writing and object lessons,[32] as it was not sanctioned by the Education Act. His justification was only partly in terms of technical education being fostered by drawing. He regarded drawing as having "the advantage of a language ... and, within a certain field it surpasses all other symbols of thought." Wilson also promised attention to lessons in agriculture in rural schools and farming districts, but he was not supported by his officials in agricultural and technical education.

The official position on curriculum changes in 1895 was a defence of the status quo because "in the nature of things, the number of children whose circumstances permit them to continue their literary education beyond the Primary school must always be small." Pupils leaving schools should be encouraged to pursue voluntarily the natural sciences, useful mechanical arts, mathematics, history, social science, or pure literature, and some form of healthy outdoor recreation.[33] All these areas had been suggested for inclusion in the school course, but were rejected as being "from their nature, impracticable, considering the environment, the children's age, the limits of time in ... schools, the money available." Enthusiasts for innovation were warned that on these matters "the red tape of departmentalism or the freaks and fashions of headmasters" should not be allowed to interfere with the existing system.[34] An expanded course could not be handled by the bulk of teachers who were "worthy and respectable men and women, of moderate capacity, of small opportunities, of limited attainments, and of humble ambitions, who do for little and do fairly well a large amount of elementary educational work." What the official position overlooked was the mounting pressure for a redefinition of the scope of primary education which was supported by overseas and inter-colonial experience, and, more embarrassingly for the Department, by some of the headteachers and school committees in centres without a grammar school where the demand for wider educational opportunity was strong enough to gain enthusiastic political action.

Early in 1895 the Department was approached by the committee of the Bundaberg Central Boys' school seeking permission to proceed with an advanced curriculum, including mathematics, book-keeping, Latin, French, English Literature, history and one or two branches of physical science, as there were no grammar school facilities in Bundaberg. This move followed a similar petition, dated 28 July 1894, and a public meeting on 12 December 1893, which supported an amendment of the Education Act of 1875 to permit the establishment of superior schools on the New South Wales pattern. A letter from the Bundaberg South School Committee on the same subject lay on the Minister's table for almost twelve months before the committee was informed that higher education was still under review. Anderson had made a very strong objection to the proposal, and with a change in the Minister on 29 March 1895 the matter rested until a petition was received from the Bowen school committee in April,[35] and then was debated in November 1895 and July 1896. Eventually a majority of Legislative Council members recognised the gap existing between the primary and grammar schools. A resolution was passed requiring the Under Secretary and General Inspector to report to parliament on the advisability and desirability of amending the Education Act of 1875 to incorporate the superior State school provisions of the New South Wales Public Instruction Act of 1880.[36]

After an extensive visit to schools in Sydney, Parramatta, and Armidale, and discussions with officials and the Headmaster of Sydney Grammar school, Under Secretary Anderson cautiously conceded the desirability of going beyond the existing primary course. With the development of mental faculties paramount in the selection of additional subjects, Anderson finally recommended the partial adoption of the New South Wales scheme and the teaching of mathematics, higher English, science, and drawing. General Inspector Ewart expressed surprise at being asked his opinion as he saw his duty confined to administrating the Act "without feeling at liberty to criticise it or to suggest its amendment." He took the view that the general taxpayer was opposed to any increase in expenditure "whatever may be the opinions and desires of individuals eager to procure at the general expense a higher education for their children." He warned against providing an excess of trained labour, and passing the control of education into the hands of the University and ambitious teachers. After a detailed examination of subjects, he proposed further work in science and mathematics, provided there was no additional expenditure. He claimed that much of the work already attempted was up to or in excess of some of the New South Wales requirements.[37]

Supporters of an extended curriculum urged the importance of education in increasing the efficiency and ability of people in a competitive world. Little had been done to improve education in Queensland "owing to having no university, and to the belief that they had a wonderfully perfect system."[38] Innovation was nedessary, despite the apparent lack of earnestness on the part of Anderson and Ewart "to see the public schools of the colony made as efficient as those in the mother colony." Discussions of the type of education which should be developed continued. The Brisbane Courier supported a broad secondary education stressing mental and character development, although it was not opposed to technical education.[39]

Confusion over the curriculum's function continued. Anderson turned to the disciplinary theory for support, while Ewart advocated changes which could be introduced without extra cost. When the Education Act was amended in 1897 both the theoretical and practical approaches emerged as part of the result, but the politically viable solution only further compromised basic differences about the function of the curriculum in general and the higher subjects in particular. The differences were exposed during the debate on the State Education Bill during 1895.[40] There was a clear demand for more than mere elementary education, but did the State have any responsibility to provide for "the special talents of a gifted few,"[41] and should higher education "be such as would educate people for trades, professions, and scientific and industrial pursuits" or should it "deal with the acquirement of mathematics, classics, and so on?"[42] The president of the Queensland Teachers' Federation at the 1896 Conference stated that the schools were not preparing for "the work-a-day world" as conservatism left pupils embedded in antiquated schedules.[43] The lack of a "stepping-stone between the primary schools and the grammar schools" warranted an extension of the elementary curriculum in the State schools.[44] Despite differences of opinion the problem did not go away, and until Anderson's retirement in 1904 neither secondary nor technical education in Queensland experienced the great transformation and the critical examination evident in Great Britain and the other colonies. This reflected the ambivalence of Anderson and Ewart, on the function of the curriculum and the distance both lagged behind contemporary thinking on curriculum development.

New Regulations issued in 1897[45] included amendments to the schedule especially in courses for arithmetic, object lessons, and drawing, and raised the standards of proficiency set for each class, again mainly in arithmetic, drawing, drill, and in the higher classes in grammar, geography, and object lessons. Temperance was included as part of object lessons on conduct and manners. The next set of revised regulations followed the passing of the State Education Act Amendment Act of 1897 which enabled the Minister to add or omit subjects of instruction in State schools. The new regulations limited the Provisional school curriculum to reading, writing, arithmetic, geography, grammar, object lessons, and needlework, and extended the courses for Classes Five and Six by adding, Euclid, algebra, an English classic, a branch of science, and a wider range of English history.[46] These changes did not clarify the general curriculum problem, and, especially as in the following years much of the teaching of the additional subjects was unsatisfactory, there was no worthwhile extension of the elementary curriculum through this innovation. Only slight progress was reported on the new courses in 1898 owing to the late arrival of the necessary textbooks, and better work was looked forward to as teachers became better acquainted with the subjects.[47] Instead of satisfactory results teachers and inspectors were soon complaining about the heavy load of work created by the new subjects.[48] The most to be claimed for the revised curriculum was that it partially reproduced the typical elementary curriculum in New South Wales.

Teachers generally resisted plans to extend the curriculum between 1898 and 1904 on the grounds of further overloading teachers and primary school pupils. From this point on, the teachers' associations tended to support the conservatism of the Department in keeping the curriculum within the narrow traditional confines of the previous thirty years. The responsibility of the primary school was "to teach the usual elementary subjects" rather than the newer additions which were confined to the Fifth and Sixth Classes, about which judgment could be delayed until the junior classes reached the higher classes.[49] Substantial reservations about the possible outcomes were expressed. As few children attended school for long enough to obtain more than a trifling amount of learning, educational paths often became blind alleys.[50] Science should be reserved for "special schools and individual attention" rather than attempted in primary school beyond the level of the facts.[51] These opinions did nothing to support the work set out in the schedule, and were echoed by some of the District Inspectors.[52] The Department opposed other changes suggested by teachers and the Queensland Education Journal in December 1899, for altering the content of school readers and for printing the readers in Queensland as suggested by the Typographical Association on several occasions dating back to 1891.

A politician unsuccessfully suggested the publication of a School Paper to follow the example of New South Wales, Victoria, and Western Australia where the local compilation contained information about Australia.[53]

A revision of the Regulations, including the programme of instruction and other schedules and undertaken by a committee of senior officers in their own time, commenced in 1900, and continued until 1904. No one particular event was responsible for this enterprise, which was, however, consistent with activities resulting in a general revamping of the primary school curriculum in most English speaking countries between 1900 and 1910. Other features of Queensland schools could be attributed to more definite events such as the South African war which led to an emphasis on patriotism in school readers and the Education Office Gazette, the erection of flag poles and the supply of Union Jacks to schools, and the singing of patriotic school songs. There was a strong interest in drill, Cadet Corps and the Teachers' Volunteer Corps.[54]

The Department confirmed the lack of sympathy with higher subjects amongst teachers and inspectors when the Minister authorised the introduction of classes in cookery at the Brisbane Technical College in June 1900.[55] No special arrangements were made for the higher subjects, but special arrangements were made in the case of practical subjects, including cookery and drill. A further concession to vocational training came in 1902 when commercial subjects, commercial correspondence and office routine, and shorthand, were added to the curriculum for the Fifth and Sixth Classes as an alternative to the science subjects.[56] A growing demand for youths trained for office work,[57] resulted in the Department yielding to pressures for practical education.

By late 1902 the higher subjects were under attack from a number of directions, amid general interest in education promoted in part by news of the English Education Bill of 1902 then before the House of Commons.[58] One conclusion, if not a rallying point, of the discussion was the need "to apply modern knowledge to the work of the modern world."[59] The Department responded favourably to teachers who applied "for permission to give practical lessons in various branches of farm and garden work" provided the new work did not "trench on the time required for the foundation subjects of elementary education."[60] While retaining the narrow elementary curriculum as a basis, the Department commenced to accept proposals for practical subjects at the same time as it expressed mounting dissatisfaction with the higher subjects of a non-vocational kind. The Courier expressed public criticism of the education system in a number of articles. After twenty seven years Queensland education had grown obsolete because it had been maintained "for all practical purposes without modification," whereas the educational systems of Germany and America contributed to strength and development. The provision of schools and teachers in a vast land was no excuse for "failure to make the system of education modern and progressive." Recent changes made on resolution of parliament or at the whim of Ministers were merely mechanical changes bearing little relation to the "strong educational movement ... taking place in other countries" of which the Minister was unaware. Efficiency was no longer assessed by maintaining "venrable departmental traditions." Further editorials criticised the curriculum as requiring suicidal perversity to regard it as satisfactory.[61]

Support for the traditional system came from a number of sources. In answer to complaints about the heavy burden of the extended curriculum District Inspector Canny advised teachers "to make their chief aim thoroughness of work in 'the three R's'. He objected to so many 'fancy branches' being included in the schedule in 1897, with a detrimental effect on achievements in elementary teaching which was regarded by him, his colleagues, and most teachers as the main function of the State schools.[62] The narrow curriculum of the Primary school was supported by several members of parliament with an attack on the overloaded schedule, although some other members proposed extra subjects because they were taught in other states.[63] Premier Philp in 1902 agreed that too many fancy subjects were taught and indispensable subjects were lost sight of as a result. He deplored attempts to alter the system because it was, in his opinion, "abreast of the times."[64]

Undoubtedly the absence of an articulated system of education from primary school to university had contributed to overloading the primary school curriculum. The Teachers' Union was supported by former teachers in parliament in its efforts to reduce the "excessive amount of work" carried by teachers. The solution was "to go back to the real foundation of primary education and commence work afresh" by establishing "a separate system of secondary education."[65] There was mounting support from the Teachers' Union and employers for a review of the education system,[66] and a narrowing of the curriculum to the three Rs was again supported in parliament.

In response to suggestions about curriculum reform the Minister, Barlow, convened the Educational Conference of 1904 attended by himself, senior officials, inspectors, and representative teachers.[67] But the Conference avoided extensive curriculum revision although sub-committee reports were prepared. These reports suggested minor modifications such as restricting Euclid and algebra to the Sixth class, introducing a school paper, eliminating parsing and analysis in the Third and Fourth classes and History in all classes up to the Fourth and introducing Civics. Some slight rearrangement of content was recommended, but as teachers were concerned with eliminating the higher subjects the elementary subjects were left substantially unaltered. The Conference shirked "the real issue"[68] by merely endorsing and updating the subjects traditionally associated with "imparting a mental culture", and making some practical concessions with cooking, woodwork and needlework.[69] Later in 1904 Minister Barlow ordered further changes in accordance with the Knibbs-Turner report in New South Wales.[70]

The revised curriculum which operated from 1905 was organised under the subjects of English, Mathematics, Nature Knowledge, Civics and Morals, Drawing, Drill and Gymnastics, Music, Needlework, and in the First class only Kindergarten. The most striking innovation was the importance given to Civics as the foundation of social behaviour and political responsibility.[71] In explaining the new schedules attention was given to practical work, correlation of the subjects of study, self-activity of pupils, and linking home and social surroundings and the influence of the school in the intellectual, moral and social development of the child. However, the Department was at pains to show that the new schedule contained nothing really new but only the best of past practices, and did little to alter the narrow three Rs dominated curriculum.[72]

The curriculum between 1876 and 1904 was subjected to few changes in content and nomenclature. It did acknowledge in 1904 the intention of meeting demands for practicality. There is no escaping the conclusion, however, that the 1904 schedule reaffirmed a narrowly conceived elementary curriculum based on an educational theory which made virtually no concessions to the demand for practical education. Increasing demands for technical, scientific and commercial training for a society becoming increasingly industrialised and less dependent on primary industry and manual labour were substantially ignored.

## Conclusion

From 1859 to 1904, and beyond, a narrow role was assigned to elementary education by the state. The curriculum remained relatively stable, in part, because there was no extension of the state's responsibility for higher, secondary and technical education. Indeed such an extension was opposed by most politicians, positively or apathetically, and by the senior officials of the Education Department.

There are some potentially significant parallels between the current exercises in both centralising and localising curriculum development and evaluation, the use of imported curriculum concepts and materials in schools and the widespread discussion of the curriculum, and similar features in the period 1859 to 1904. The control of the curriculum was maintained firmly by the central administration, although some local variations were not prohibited. Major imported curriculum projects were represented by the adoption of series of readers and textbooks from overseas. There was some public discussion of the curriculum, some pressure for reform was applied, and spasmodic attempts were made to evaluate the success of the curriculum in both the professional and political spheres, as there are today and with much the same sort of minimal impact on what is produced and what is taught. In the colonial period curriculum development and evaluation were largely determined by administrators, operating within a broader political and intellectual context which accepted and fostered the narrowly conceived elementary curriculum based on the development of mental faculties. Although the dominance of the administrators did not go unchallenged, curriculum development and evaluation remained largel intuitive, idiosyncratic, imitative, and certainly without objective criteria. The administrators followed the inherited system, their own predilections and prejudices, and they moved only very reluctantly to adjust to changes being made to established curriculum models overseas and in other colonies.

1. Letterbook, 18 May 1860, and The Education Act of 1860, clause VI.
2. Courier, 14 April 1860.
3. Letterbook, 1870, 70/70, 5 April 1870.
4. Report 1872, pp.10-11.
5. Royal Commission Report 1874, pp.277-79;
   Queensland Parliamentary Debates 1874, vol.16, p.395.

6. *Royal Commission Report* 1874. Minutes of Evidence, pp.17,19,25.
7. *Report* 1876, pp.22-24.
8. *Report* 1877, p.29.
9. *Queensland Parliamentary Debates* 1878, vol.26, p.894.
10. *ibid.*, 1880, vol.33, pp.805-06.
11. *Department Various* 1880, 14 August 1880, 15 September 1880.
12. *Report* 1887, p.45.
13. *Report* 1904, p.27.
14. *Report* 1885, p.4.
15. *ibid.*, p.68.
16. *Report* 1886, p.51.
17. *Queensland Parliamentary Debates* 1886, vol.50, pp.1430-33.
18. *ibid.*, 1887 vol.53, pp.1504-05.
19. *Report* 1888, p.78
20. *Report* 1889, pp.23 & 28.
21. *Executive Council Minute Book*, 1890, 90/88; 1891, 91/53
22. *Queensland Government Gazette* 1891, 6 July 1891, pp.761-83.
23. *Report* 1889, p.67.
24. *ibid.*, p.75.
25. *ibid.*, 1890, p.8.
26. *ibid.*, 1889, p.69.
27. *ibid.*, 1890, p.6.
28. *ibid.*, 1889, p.77.
29. *ibid.*, 1892, pp.6, 12.
30. *ibid.*, 1893, p.55.
31. *ibid.*, 1892, p.56.
32. *ibid.*, 1893, p.6.
33. *ibid.*, 1895, p.6
34. *ibid.*, p.49.
35. *Department Various* 1895-1898 - Papers relating to the establishment of Superior Schools.
36. *Journal Legislative Council* 1896, vol.46, pt.ii, pp.1063-1097.
37. *ibid.*, pp.1074-1079.
38. *Queensland Parliamentary Debates* 1896, vol.76, p.1509.
39. *Courier*, 16 August 1897.
40. *Queensland Parliamentary Debates* 1895, vol.74, pp.1517-1531, 1676-1688.
41. *ibid.*, p.1529.
42. *ibid.*, p.1682.
43. *Courier*, January 9 1896.
44. *Queensland Parliamentary Debates* 1896, vol.75, p.385.
45. *Queensland Government Gazette* 1897, 13 March 1897, pp.791-803.
46. *ibid.*, 1898, 21 May 1898, pp.1408-1409.
47. *Report* 1898, p.52.
48. *ibid.*, p.53.
49. *Report* 1899, p.53.
50. *ibid.*, p.55
51. *ibid.*, p.56.
52. *ibid.*, 1899, p.63.
53. *Queensland Parliamentary Debates* 1899, vol.83, pp.1368-1369.
54. *Report* 1900, pp.56,62,7.
55. *Report* 1901, p.7.
56. *Queensland Government Gazette* 1902, 19 July 1902, p.170.
57. *Courier*, 11 January 1902.
58. *ibid.*, 27 October 1902, 8 November 1902, 15 November 1902.
59. *ibid.*, 8 November 1902.
60. *Report* 1902, p.7.
61. *Courier*, 6 December 1902, 10 December 1902, 12 December 1902, 13 December 1902.
62. *Report* 1902, p.73.
63. *Queensland Parliamentary Debates* 1901, vol.88, p.2193.

64. ibid., 1902, vol.90, p.1484.
65. ibid., pp.1563-65.
66. Courier, 7 January 1903.
67. Queensland Parliamentary Papers 1904, Proceedings of the Educational Conference, 25 to 29 January 1904, pp.277-300.
68. Courier, 30 January 1904.
69. Queensland Parliamentary Papers 1904, pp.289-90.
70. Report 1904, p.23.
71. Queensland Government Gazette 1904, 15 October 1904, pp.821-26.
72. Report 1903, pp.6, 23-7.

School of Education
Gippsland Institute of Advanced Education
Churchill, Victoria    3842

# RESPONSIVE EVALUATION OF A PRIMARY SCHOOLS SIMULATION PROJECT (E.R.D.C.)

C. Marsh and K. Treharne

## INTRODUCTION

The frequently cited dichotomy between achievement testing using strictly controlled experimental conditions and informal observations and opinions by panels of experts has become blurred over recent years as evaluation specialists have attempted to select out advantages from both approaches. (Stake (1975) Wynn (1976) ).

Evaluation specialists are tending to incorporate informal observations, rating scales and interviews into their research design in addition to standardised tests, even though the penalty for this broadening of instruments often requires smaller samples to be studied. (MacDonald (1971), Hamilton (1976), Wynn (1976)). Data is being sought which not only attempts to measure the implementation of intended goals, but also unintended goals. Although there are critics of the goal-free approaches to evaluation (Alkin, 1972) there are advantages in broadening evaluation studies if they are to portray what actually happens (Scriven (1972) Stake (1967)). Also, recent attempts by evaluation specialists have focussed attention on process as well as product variables. The inclusion of process variables provides valuable feedback to developers during the vital developmental periods, without removing the need for some measurement of outcomes exhibited by the final product.

## RESPONSIVE EVALUATION

One such evaluation approach which incorporates these changes is the responsive evaluation model developed by Robert Stake (1975), which emphasises observations and reactions between evaluators and programme participants. This type of evaluation is characterised by an orientation toward programme activities rather than programme intents; direct responses to audience requirements for information and a consideration of the different value perspectives of various personnel involved in the programme (Stake 1975, p10).

To undertake a responsive evaluation requires collecting data from a variety of sources and communicating and receiving information from a number of persons. The evaluator often starts with a plan about data sources to be used and major issues to be explored. These are then discussed with audience personnel. Their values and preferences (responses) are valuable for the evaluator in extending the major issues. With the assistance of the personnel involved, the evaluator will periodically develop portrayals of the processes involved in implementing the programme. Further responses by audience personnel to these portrayals will guide the evaluator in developing subsequent data collecting instruments.

Stake (1975, p16) maintains that it is crucial for the evaluator to get his data in sufficient amount from a wide range of independent and credible sources. To this end, the matrix he developed in an earlier paper (Stake, (1967), is still appropriate in providing a range of data sources. (Fig. 1).

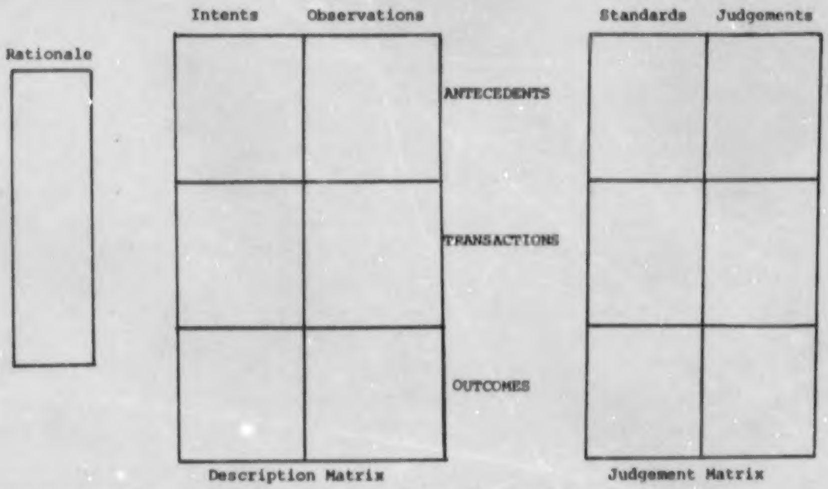


FIG. 1 (Adapted from Stake, R.E. The Countenance of Educational Evaluation Teachers College Record, LXVIII, 1967, 523-540).

It is up to the evaluator to judiciously select from this matrix - collecting data - responding to it - communicating it. Stake (1975, p19) sees this as a series of events or interactions between the evaluator and audience personnel which occur continuously and spontaneously and not necessarily in any linear sequence (Fig. 2).

Talk with clients, program staff, audiences

Identify program scope

Overview program activities

Discover purposes concerns

Conceptualize issues, problems

Identify data needs, re, issues

Select observers, judges, instruments, if any

Observe designated antecedents, transactions and outcomes

Thematize; prepare portrayals, case studies

Validate confirm, attempt to disconfirm

Winnow, format for audience use

Assemble formal reports, if any

FIG. 2 PROMINENT EVENTS IN A RESPONSIVE SITUATION

(Adapted from Stake, R.E. Program Evaluation, Particularly Responsive Evaluation, Occasional Paper Series No. 5, p19).

Rippey's concept of transactional evaluation (1973 p3) has a lot in common with responsive evaluation in that the emphasis is upon collecting data about changing role relationships of participants in a programme. To elicit information about their expectations, doubts and apprehensions, communication between the evaluator and participants is essential. This information is then fed back into the system to enable the participants to clarify their roles and to ensure more productive interactions during the operation of the programme. However, responsive evaluation does not go to the next stage of transactional evaluation in which conflict issues are explored by setting up proponents and opponents to develop and implement their own evaluation programmes.

Responsive evaluation is most appropriate in situations when developers need feedback about ongoing perceptions of participants, emerging successes and problems. However, its value is not limited to these formative elements of evaluation because summative evaluation demands are also met by its ability to provide a comprehensive portrayal of the strengths and weaknesses of a programme.

It is readily apparent that this evaluation approach is admirably suited to any simulation-based programme which focusses upon affective growth, in which interactions between participants are of considerable importance and where outcomes are largely unknown, apart from general goals. Evaluation studies on simulations to date have been very meagre and the evidence has been equivocal, at best (Gordon 1972) Livingstone (1973)). Teacher education simulations have not been the subject of many research studies but standardised tests have been used by some researchers and produced evidence that positive transfer from simulations to classroom practice does occur in certain conditions (Cruickshank and Broadbent (1968)) and that it is possible to use predictive instruments for slecting successful solvers of classroom problems (Haefele (1976)). The use of 'illuninative' evaluation procedures, to obtain data about teacher education simulation programmes, does not appear in the current research literature.

# A RESPONSIVE EVALUATION MODEL FOR A TEACHER EDUCATION SIMULATION

The Primary Schools Simulation Project, initiated under an ERDC grant in 1976, has resulted in the preparation of eight teaching package units which reflect the general theme of a Challenge of Change. Three units were completed by the end of 1976 and deal with such issues as Changing Aims, Changing Relationships and Changing Environments. The remaining five units are almost completed and focus upon such issues as changes in resource management and organisational patterns, and interpersonal relations between parents, students, and fellow staff members. The discussion situations are presented via small group workshops and role playing sessions, after which it is recommended that follow-up activities should occur in practice teaching periods. The sequence has as its conceptual base, the progression from an elaboration of an issue to the taking of a stand in a non-threatening situation to the implementation and observation of the orientation in actual classroom situations. (Fig 3).

**SEQUENCE WITHIN EACH UNIT**

1. Introduce major concept-theme using written and multi-media stimulus examples.
2. Provide workshop training exercises in which small-group discussion techniques are used to consider descriptive, analytic and prescriptive issues related to the major concept.
3. Provide role-playing examples which simulate typical classroom decision situations, using combinations of stimulators such as filmed incidents, written incidents, role-play cards, audio tapes.
4. Provide follow-up activities based upon the major concept, to be undertaken in the schools during unit or block practice periods.

FIG. 3.

An evaluation model was designed to provide information about these teaching packages, especially the first three units which are currently being trialled in teacher education programmes in three tertiary institutions. Each of these units focusses upon an aspect of educational change, applicable to teaching in an Australian primary school. Ultimately, the goal of each unit is to produce attitudinal changes among the student participants but the magnitude of these are extrememly difficult to gauge. On a more realistic level, the goals for each unit are :

(a) to enable students to widen their range of attitudes and values toward specific educational issues included in the unit.

(b) to improve students' ability to base their judgements about issues upon rational thinking processes.

Following Stake (1975), the first step in the development of a responsive evaluation model is to outline the potential audiences and the range of data that they will require (Fig. 4). In this instance there appeared to be a number of potential audiences, including pre-service and inservice teacher educators, state department inservice educators, preservice teachers, inservice teachers, senior education administrators and parents. The task of the evaluator becomes more complex if wider audiences are involved. Taking dimension A (Fig 4) where an instructor uses a unit with a group of students over a semester, the audiences involved are only one teacher educator and twenty preservice teachers. If a wider evaluation programme is envisaged to encompass the effects of using the unit upon a total teacher education institution (dimension D, Fig 4), then clearly additional audiences are involved including numerous teacher educators, senior administrators and preservice teachers not involved in the unit. The evaluation task becomes even more complex if the study is widened to encompass inservice teaching within a state system of education (dimension F, Fig 4). Initial decisions therefore are required between the evaluator and the funding body about the dimension of the evaluation programme in terms of audience priorities and cost benefits.

An early task for the evaluator is to interact with the audience participants, observing behaviour, reacting to their proposals and making suggestions. These preliminary activities are necessary in drawing up a matrix of targets which might be used to evaluate the unit. It is up to the evaluator to identify important activities, to conceptualise problems raised and to provide an overall framework from which selections and decisions can be made by the funding body.

As an example, a range of evaluative activities are listed in Fig. 5 which focus upon the behaviours of just the instructor for the unit and student participants. These are categorised into antecedent, transactional and outcome behaviours and the instruments listed for each category are indicated as minimal examples. It would be up to the funding body to make judgements about which categories they would want included, and any additional instruments which might be needed.

| | DIMENSION A | | DIMENSION B | DIMENSION C | DIMENSION D | DIMENSION E | DIMENSION F | DIMENSION G |
|---|---|---|---|---|---|---|---|---|
| R.E. Evaluators | Instructors | Preservice Participants in the unit | Other Teacher Educators | Preservice Teachers not involved in the unit | Administrators in Teacher Edn. Programme | Inservice Teachers | State Dept. of Edn. Personnel | General Public |
| ANTECEDENTS | | | | | | | | |
| TRANSACTIONS | | | | | | | | |
| OUTCOMES | | | | | | | | |

FIG. 4 MATRIX OF DIMENSIONS OF ANALYSIS AVAILABLE WITH RESPONSIVE EVALUATION TECHNIQUES

Once these have been selected, the evaluation can provide the relevant data required, and in particular, indicate degrees of congruency between intentions and outcomes and contingencies between variables in the matrix (Stake 1969, p 17). For example, student attitudes about issues in the unit as expressed by the instructor prior to the first workshops, can be compared to the stated attitudes expressed by students in their participant inventory sheets. The degree of congruence between the two sources of data can be analyzed and communicated back to the instructor and participants.

The contingencies between the variables are more difficult to analyze beyond statements of description. As indicated above, it is unlikely that evaluative evidence could provide sufficient data to indicate that attitude changes were caused by participation in the unit. However, it may be possible to use some of the transactions and outcome data to describe some apparent shifts, even if they were simply a widening of value choices and a willingness to consider choices rationally before making a decision.

The advantage of the responsive evaluation approach is that the monitoring of behaviour by the evaluator is continuously being relayed to the instructor and participants so that they can make necessary adjustments. At the formative stage this is most useful, but also at the completion of the unit, the responsive data provided from the summative evaluation instruments enables feedback to instructors and students which may assist them in undertaking future units and related activities.

DIMENSION A

| INSTRUCTOR | | STUDENT PARTICIPANTS | |
|---|---|---|---|
| ANTECEDENTS | INSTRUMENTS USED TO COLLECT DATA ABOUT : | ANTECEDENTS | INSTRUMENTS USED TO COLLECT DATA ABOUT : |
| Teaching experience | Questionnaire | Practice Teaching | Questionnaire |
| Teaching Philosophy | Questionnaire and interview | Previous Employment | Questionnaire |
| Attitudes about issues in Unit | Questionnaire & interview | Early Education | Questionnaire |
| Relationships with Students | Questionnaire & interview | Atticudes about issues in Unit | Semantic Differential Pretest<br>Observations by Instructor |
| | | Formal courses in education | Questionnaire |
| TRANSACTIONS<br>Supervising small groups | Instructor Diary<br>Independent Observer | TRANSACTIONS<br>Involvement in issues | Participant Inventory<br>Observations by Instructor, (Instructor Diary, Observation sheet)<br>Independent Observer |
| Reacting to issues raised | " " | Expressing opinions | " " |
| Maintaining discussion | " " | Taking a stand on an issue | " " |
| Encouraging role plays | " " | Expressing empathy | " " |
| Supporting discussion of role plays | " " | Actively involved in role playing | " " |
| OUTCOMES<br>Attitudes about the issues | Interview | OUTCOMES<br>Articulating about issues | Individual and group interviews |
| Attitude about student performance | Interview | Expressing attitudes about issues | Instructor's diary<br>Interviews<br>Semantic Differential Post-Test |
| Attitudes about the workshops and role plays | Interview | Analyzing related issues | Written critical incidents example |
| Attitudes about the effectiveness of the unit. | Interview | | |

FIG. 5 MATRIX OF DATA TO BE COLLECTED
BY THE EVALUATOR

## TRIALLING ACTIVITIES WITH THE RESPONSIVE EVALUATION MODEL

The responsive evaluation model used with the three trial units was limited to the immediate dimension A (Fig. 4) of individual instructors and their respective classes. Consequently the instruments used were only those pertaining to instructors and participants, as indicated in Figure 5 and interactions with the evaluator were therefore limited to these persons. The model was also limited in these three trials in that students and instructors were informed that results from the evaluation instruments would not be used to award grades for the unit. The emphasis was upon receiving information about the usefulness of the components within the unit and the evaluation procedures, rather than assessing student performance.

A number of issues appeared early in the trialling which provided opportunities for useful interaction. The overall sequence for each of the units was welcomed by both instructors and participants. The participant inventory details and the instructors' diary notes both indicated that the sequence of workshops through to role plays was logical and quite successful. In fact the discussion of basic issues in the workshops seemed to pave the way for ready acceptance of role playing activities. One difficulty involved maintaining the integrity of the workshop materials. One instructor varied the materials used, deleting a number of the recommended readings and inserting alternative activities. The evaluator was able to reach a consensus with the instructor on this occasion, without reducing the latter's enthusiasm for using the materials in the unit. However, this incident raises the issue of whether the unit material is the determining variable or whether it is the personality of the instructor.

The evaluation instruments were able to provide some information about a widening of attitudes even though it was not possible to confirm that participant attitudes had actually changed. The semantic differential instruments were not developed until late in the trialling and so these could not be used to measure attitudinal change. Early attempts at developing appropriate concepts and scales proved to be very time-consuming, and largely unsuccessful. The final version was based upon four of the concepts and ten of the bipolar scales developed and validated by Elsworth and Coulter (1976). However, the participant inventory, diary notes and interviews provided a number of clues about variations in student behaviour. Common remarks included
"emotional insights from the roleplays helped to draw out contradictions"
"role plays brought out the quiet students"
"most became less dogmatic".

Comments received early in the trialling indicated that the evaluation dimension as depicted in Fig. 5 was too narrow. Students participating in the unit indicated that the materials should have been included in educational psychology and sociology courses as well as teaching practice courses. The instructors also reiterated the view that it would have been desirable to bring in staff members from these disciplines. It seems important that opinions of a wider number of teacher educators are needed in future trialling as the units have application beyond teaching practice courses. In these three instances, the temporary nature of the trialling process was sufficient for the participants to acquiese but it is a policy decision which will need to be considered.

Finally, a number of revisions were suggested about the evaluation instruments. The observation sheet used by the instructors apparently contained misleading and ambiguous questions and will need to be reworked. Both structured and unstructured interviews were used at the completion of the unit with instructors. The structured interviews which required indepth answers about the workshops of the unit produced more specific information and seem worth refining still further.

## SUMMARY

A responsive evaluation approach enjoys the benefit of comprehensive portrayals of information to audiences, even if a debit is a reduction in precise measurement. It was pointed out that simulation programmes are not conducive to measurement of students' achievement and that the responsive evaluation approach is helpful in capturing the spirit of the behaviour of participants.

The model has been used during the trialling of three simulation units. The data collected has been especially helpful at the formative stage in the development of these units and it is envisaged that the same techniques will be valuable information for instructors to use when completed versions become available. From the data obtained in the trials, the units do appear to broaden student attitudes towards specific educational issues. More sophisticated instruments used in subsequent trials may be able to uncover more direct relationships between the materials and attitudinal changes.

| | |
|---|---|
| Cruickshank, D.R. & Broadbent, F.W. | The Simulation and Analysis of Beginning Teachers, Final Report, Project 5 - 0798, Washington, D.C: U.S. Department of Health, Education & Welfare, 1963. |
| Elsworth G.R. and Coulter F. | "The Development, Reliability and Validity of a Scale to Measure Professional Self Perception in Student Teachers" Technical Report to ERDC, August 1976. |
| Gordon, A.K. | Games for Growth, Chicago: SRA, 1972 |
| Haefele, D.T. | An Investigation of Factors Associated with a Teaching Problem Simulation Simulation and Games VII, 3 September 1976, 311-319. |
| Hamilton, D.F. (Ed) | Reader in Illuminative Evaluation Berkeley, California: McCutchan, 1976. |
| Harlen, W. | "Change and Development in Evaluation Strategy" in Tawney, D. (ed) Curriculum Evaluation Today: Trends and Implications London: Macmillan Education, 1976. |
| Livingstone, S.A. & Stoll, C.S. | Simulation Games: An Introduction for the Social Studies Teacher New York: The Free Press, 1973. |
| MacDonald, B. | The Evaluation of the Humanities Curriculum Project: a holistic approach Theory into Practice, X,3, 163-167. |
| Rippey, R.M. | "What is Transactional Evaluation" in Rippey, R.M. (ed) Studies in Transactional Evaluation Belmont, California McCutchan, 1973. |
| Scriven, M. | "Goal Free Evaluation" in Rippey, R.M. (ed) Studies in Transactional Evaluation Belmont, California McCutchan, 1973. |
| Stake, R.E. | The Countenance of Educational Evaluation Teachers College Record LXVIII, 1967, 523-540. |
| Stake, R.E. | "Program Evaluation, Particularly Responsive Evaluation", Occasional Paper Series No. 5, Evaluation Centre, College of Education, Western Michigan University, Kalamazoo, Michigan 49008, 1975. |

# The probable limits of early intervention programmes for the substantially intellectually retarded

At present in New Zealand there are quite strong pressures being generated to begin widespread intervention programmes with very young intellectually handicapped children (e.g. Stratton, 1977). New Zealand often follows overseas trends in non-critical and inappropriate ways, and this tendency seems to be being reflected in the present instance. It is as if the mistakes of Head-Start and derivative programmes will need to be repeated locally with practically no reference to the lessons gained from the expenditure of ten billion dollars.

The arguments in favour of intervention moves are tenuous, hopeful and over-generous in their interpretation of the efficacy of present and past programmes. There is, for example, little evidence to point to the specific procedures that such programmes should be based upon, and little regard has been paid to the limits of effectiveness of such programmes given even ideal conditions.

The general point of view seems to be that because early intervention programmes have worked with some forms of handicap it should work for most if not all. But this is more an article of faith than a valid conclusion drawn from the results of studies in the area. For example, measure taken with children's sensory handicaps are sometimes referred to by early interventionists. It has been clearly demonstrated that fitting a hearing aid to very young deaf infants gives them better hearing in later life than they would have had if the aid had been delayed by even a couple of years. More recently it has been found that some cases of congenital squint can be cured in the neonate by non-surgical procedures (Blakemore, 1974). The child is shown an irregular gingham pattern on a curved surface, at two feet distance, at intervals, for a total of about 30 minutes per day, and the condition disappears. The hypotheses used to explain this are fairly complex.

Similarly, results of intervention programmes with culturally depressed groups can be invoked as part of the argument as are studies on manipulation of sensory input arising from the work of Bexton, Heron and Scott (1954) or the later studies of Rosenzweig, Diamond, Krech and Bennett (1964) on the effect of sensory deprivation in animals on the development of the central nervous system (CNS). In fact, when one reviews the pot-pouri of evidence that the early interventionists provide as justification for their proposals one is forced to conclude that educators (in the widest sense) have not taken to heart the lesson spelled out by Dewey when he pointed out the common fallacy in education of what he called either/or thinking. In the present case it is reflected in a failure to accept the fact that the kinds of studies cited apply to very specific handicaps which still have only limited results. Let us now consider points in the case of the early interventionists in more detail.

In early intervention programmes with deaf children the condition being treated is specific and so is the treatment. Indications are that the treatment is more likely to work where the condition is not the result of CNS impairment, but due to a failure of some part of the mechanical apparatus instead. Moreover this procedure does not restore the function to normal but only *reduces the final deficit*. The same is true of the early treatment of congenital squint, in that the treatment is very specific and was arrived at as an accidental bi-product of some research with animals.

Along other lines it has been argued that since sensory isolation produces dysfunction in the CNS then surely perceptual enrichment will enhance whatever abilities may be there, and results from studies by the Rosenzweig team are invoked to support this. Almost complete *sensory isolation* is quite a different thing, however, from *socio-cultural deprivation* and is much more extreme than even the most severe clinical cases of deprivation reported in the literature, such as the more recent Czech report (Koluchova, 1972, 1976). About such cases Jensen (1969) has correctly observed that return to a normal environment whether middle-class or otherwise produces "... a rapid, marked and permanent improvement..." in the functioning of the children. Whether it improves ultimately to the point it might have been without the deprivation is something we just do not know at present. In cases where the changes do not follow Jensen's picture we assume that there was initial damage to the CNS, such as in the case of unresponsive feral children (Dennis, 1941).

The significance of the findings of the Rosenzweig studies are easily misinterpreted or misused. Summarising briefly they had three groups of rats, one group was perceptually restricted another (the control) had conditions that are typically provided for laboratory rats, while the third (enriched) group had a situation that was more varied than that usually provided for them. Histological study of the brains of the three groups showed the enriched group to have slightly heavier brains than the controls with a slightly higher concentration of neurotransmitters and their derivatives, while the control group was substantially greater than the deprived group in these respects. Rosenzweig's team concluded that environmental complexity is a factor in CNS development and this has been borne out in a number of subsequent studies (See Valverde and Ruiz-Marcos, 1970, for example).

The problem here is to identify the norm in order to see whether artificial enrichment can enhance natural potential, or instead, whether isolation subtracts from natural potential. Since typical laboratory rats manifestly do not enjoy the range of sensory experience of rats in their natural habitat we can conclude that the enriched condition was the closest of the three conditions to natural. Thus while we can conclude that perceptual restriction reduces natural potential we cannot say that the same studies show that enrichment enhances it. There may be a continuum of effect, but the decrease in the differences from the restricted to typical to enriched suggests that an asymptotic situation is being approached which would point to little if any significant improvement beyond the laboratory enriched condition.

Turning to studies on the group of children who are socio-culturally depressed, we see in summaries like Bronfenbrenner's (1974) confirmation of the view that the amount of benefit demonstrated by programmes like Head Start is related to the degree of recency of the programme at the time at which the subjects are assessed, with a deline associated directly with the distance in time from the termination of the project. In fact, only where the programme has gone far beyond the child in the school and looked at the family situation as a whole has there been evidence of lasting effects.

For most of the children in the Head Start and kindred programmes we assume generally that there is no substantial or significant impairment of the CNS, and also that there is a period of non-debilitated learning. The significance for recovery or improvement of children who have had a period of unimpaired learning experience has often been noted, since it gives them a considerable advantage over those whose learning experience has always been markedly handicapped. Thus it seems appropriate to regard the changes in those receiving intervention because of depressed socio-cultural environments as the upper limit of what we might expect for children who are already substantially retarded at birth, but there is no positive evidence to suggest that all or most are likely to maintain even that modest goal.

Despite this picture we get claims emanating from the early interventionists which ignore the evidence to varying degrees. Perhaps the most extreme position in this respect is taken by Doman, who in response to a question about how much improvement one could expect in substantially retarded children treated strictly according to his programme, replied "... all the way ..." in about a third of the cases, meaning presumably that those cases would improve sufficiently to come within the normal range. In fact no-one has shown me a subnormal child on that or any other programme who has improved to remotely near that level, neither have I had any verified reports of such cases in New Zealand or elsewhere. Thus there is an assumption, explicit in Doman's case, at least, and implicit in others that appropriate early intervention will compensate for most if not all of the many and varied disadvantages of impaired CNS equipment, and indeed may even repair it.

Others who have worked extensively at habilitating retardates such as Kephart (1963) and more particularly Cratty (1972) make very modest claims about what might be achieved, and they do not assume that behavioural manipulation of children will restore functions to their initial levels. The experience of therapists who have begun to use electromyographic feedback to treat people with CNS damage would also discount the suggestion that they may be promoting the repair of the CNS to something like its original state. (Johnson and Garton, 1973). This is not the place to go into details about the effects of training on the damaged CNS. It is probably sufficient to remind you that there is no regeneration of damaged tissue in the CNS, and that any subtraction is permanent. Where lobes have apparently paired functions (such as the frontals) unilateral injury can be accommodated to a limited extent by the unimpaired partner, but it is still possible to see signs of a decline overall in the quality of performance. (Luria, 1974, gives a useful up-to-date summary of this evidence.)

A very recent study by Ramey and Smith (1977) does tease out some of the considerations involved and gives a good insight into the limitations of even very broad early intervention programmes based on current knowledge and ideas. They took a group of children who were at high risk for mental retardation from severe socio-cultural disadvantage. At age 6 months the children were assessed on various tasks and then divided into two groups, one to be a matched control and the other to receive a fairly comprehensive kind of intervention programme involving social welfare, pre-school and health facilities for the whole family. A check when the children were 18 months old showed a marked and interesting difference, which is probably best reflected in the results for the Bayley Mental Development Index, as follows:

| Age | 6 months | 18 months | |
|---|---|---|---|
| Experimental Mean MDI | 106 | 102 | |
| Control Mean MDI | 104 | 86 | $p<0.02$ |

Note that the intervention programme is working, but rather than increasing the initial developmental status of the experimental group it is simply preventing it from declining at the same alarming rate as the control group.

Quite possibly what we are seeing in the effect for the control group is the underfunctioning of the potential that is there. Hopefully it might be reached again given the proper conditions, at least the possibility is suggested by the experience of Feuerstein and Kraisilowsky (1972) with a group of deprived adolescents who were apparently rehabilitated. In that group also, as Clarke and Clarke (1977) point out, the causes of their retardation were psycho-social and not physical. They were congenitally normal and also had the benefit of at least some unimpaired learning experience. On the other hand children who have a moderate or greater degree of retardation generally have a congenitally lower level over capacity as well as the additional disadvantage of no unimpaired learning experience. Consideration of the principal factors in the Smith and Ramey study suggest that if the population could have access to the kind of facilities already enjoyed by at least the upper 60% of populations in the more developed western communities then there would be little need for intensive kinds of intervention to overcome psycho-socially induced retardation.

In conclusion then, when we analyse the evidence about early intervention instead of the claims for it, a very consistent picture emerges (regardless of the type or source of evidence) which leads to the following observations.

1. In evaluating the potential for success of early behavioural treatment programmes it is necessary to identify between the types of mechanisms involved. When this is done it becomes evident that where such programmes have

worked best (with disorders having a physiological basis) the critical element in the condition has been muscular and not neurological.

2. That where programmes are aimed at compensating for the effects of a limited socio-cultural environment these are most effective when the activities go beyond the immediate target individuals.

3. That there are upper limits to the effects of optimising the environment for neurological development and in only some limited types of cases will this be as high even as the congenital level of ability.

4. Thus in the sense that these programmes only reduce a further decline in the amount of handicap we can conclude that they are fundamentally preventive. They are not ameliorative, however, in the sense that they do not enhance the limited initial capacity.

A full awareness of these observations and their implications could be of widespread help. Hopefully this would protect retarded persons from the excess practices of some current programmes. It should also benefit their families as a whole by reducing the pressure on them, by minimising the demands on their time and resources spent in seeking out and slavishly following the programmes that they have been encouraged to believe will make their child almost normal or even entirely so.

Present knowledge would suggest that supportive counselling for the parents (including information giving) some extra paediatric services, provision of the proper day care and short stay centres for relief, plus advice on how to keep up a good but not excessive level of stimulation for the child in the home would achieve at least as much as the frenetic programmes one can find at present.

## REFERENCES

Bexton, W., Heron, W. & Scott, T. (1954) Effects of decreased variation in the sensory environment. Canad. J. Psychol., 8, 70-76.

Blakemore, C.B. (1974) Effects of visual experience on the developing brain. Mod. Probl. Paediat., 13, 229-33.

Bronfenbrenner, U. (1974) A report on longitudinal evaluations of pre-school programs., vol.2. Is early intervention effective? Washington D.C. DHEW Pub. no (OHD) 74-25.

Clarke, A. & Clarke, A. (1977) Prospects for prevention and amelioration of mental retardation. Amer. J. Ment. Defic. 81, 523-33.

Cratty, B.J. (1972) Physical Expressions of Intelligence N.J.: Prentice Hall.

Dennis, W. (1941) The significance of feral man. Amer. J. Psychol. 425-32.

Feuerstein, R., & Krasilowsky, D. (1972) Interventional strategies for the significant modification of cognitive functioning in the disadvantaged adolescent. J. Amer. Acad. Child Psychiat. 11, 572-82.

Jensen, A.R. (1969) How much can we boost IQ and scholastic achievement. Harv. Educ. Rev., 39, 1-123.

Johnson, H. & Garton, W. (1973) Muscle re-education in hemiplegia by use of electromyographic device. Arch. Phys. Med. Rehabil. 54, 320-25.

Koluchova, J. (1972) Severe deprivation of twins: a case study. J. Child Psychol. Psychiat. 13, 107-14.

Koluchova, J. (1976) Further development of twins after severe deprivation. In Clarke, A. & Clarke, A. (eds) Early Experience. London: Open Books.

Kephart (1963) The brain injured child in the classroom.

Luria, A.R. (1973) The working brain. London: Penguin.

Ramey, C. & Smith, B. (1977) Assessing intellectual consequences of early intervention with high risk infants. Amer. J. Ment. Defic. 81, 318-24.

Rosenzweig, M., Diamond, M., Kriech, E. & Bennett, L. (1964) Experience and brain plasticity. Science, 146, 310-19.

Stratton, E. (1977) The dawn start project. Ment. Handicap N.Z. 2, 14-35.

Valverde, F. & Ruiz-Marcos, A. (1970) The effects of sensory deprivation on dendritic spines. In Young, F. and Lindsley, D. (eds) Early experience and visual information processing in perceptual and reading disorders. Washington: N.A.S.

TRANS 6/7

# THE TRANSITION OF STUDENTS
# FROM
# PRIMARY TO SECONDARY SCHOOL

Report of a preliminary study by
Geoffrey R. Trebilco (Rusden State College)
Elaine P. Atkinson (Monash University)
John M. Atkinson (Rusden State College)

Paper presented at the Annual Conference of the Australian
Association for Research in Education, Canberra,
November, 1977

Special acknowledgement is made of the assistance given by

Mavis Grierson, Rusden State College
Barry Milne, Computer Centre, Monash University

## INTRODUCTION AND RATIONALE

"The transfer usually involves a change of school, a parting with at least some companions and forming new relationships, a greater variety of teachers, becoming again 'infants' in a new setting instead of members of a top class, an introduction to new subjects and teaching methods and exacting home tasks. These are heavy demands on any child, and nothing in the transfer arrangements themselves should increase the strain of the changeover or produce in the child a feeling of crisis."

So wrote the author of a report of the Scottish Advisory Council on Secondary Education (Scottish Education Department, 1947, p.192) a generation ago.

The transition from primary to secondary education has long been a matter of concern. The Ramsay Report (Vic. 1960) noted certain difficulties faced by pupils as they move from Grade 6 to Grade 7 and implied that many children have transfer problems. Both the Wyndham Report (N.S.W., 1959) and the Dettman Report (W.A., 1969) made recommendations aimed at smoothing the transition from primary to secondary school.

The general concern over transition is no doubt heightened by the fact that the transfer occurs at what is an age of instability for most youngsters anyway. A middle school in the United States described its typical client thus:
"a preadolescent emerging from childhood in slow stages, sometimes awkward and insecure, sometimes facile and adept, frequently concerned with self assessment, often amazed by newly-developing powers, constantly in need of appropriate opportunities for exploration and venture, sometimes capable of adult behaviour and responses, and frequently in need of opportunities for trial-and-error in situations where error is acceptable." (Encyclopedia of Education, 1971, quoting from Educational Facilities Laboratories, 1962, p. 19.)

In some countries (e.g. the United States, New Zealand and the United Kingdom) intermediate or middle schools which embrace Grades 5 or 6 to 7 or 8 have been introduced, partly to overcome difficulties of transition (as well as for other reasons not directly relevant to the present discussion). However, both the Wyndham Report and the Dettman Report rejected this option, pointing out that it replaces one transition with two.

In Australia the approach has been to retain the two stage system[1] of schooling involving a transfer at about age 12, whilst attempting to alleviate problems presumed to be associated with the transfer.

Records from the Curriculum & Research Branch of the Education Department of Victoria show that a series of meetings about "the transition problem" date back to 1966. Subsequent meetings have been held in many Victorian centres as well as in the metropolitan area. The initial meetings followed a report of responses to a questionnaire administered to Grade 7 pupils several weeks after they commenced secondary school (Education Department of Victoria, Curriculum & Research Branch, 1966). The responses of 318 pupils sampled from a total of 2,300 returns from 16 schools (7 high schools, 9 technical schools) were analysed. One section of the questionnaire asked students to indicate whether each of twenty six situations worried them a great deal, a little, or not at all. For analysis, situations were classified as being "worries that could be overcome during the first week of school" (e.g. I do not know the teachers who teach me; I do not know where to buy my lunch), "worries caused by external circumstances" (e.g. I am separated from my best friend), and "worries caused by personal inadequacies" (e.g. I might not understand new work; I am shy). The report made a number of suggestions on the basis of the results of the survey, mostly relating to orientation needs. Many secondary schools now have special procedures or organisational structures designed to "ease" incoming Grade 7 pupils into the secondary school. Some schools have gone so far as to incorporate a special "transition school" for Grade 7.

---

[1] Middle schools have been proposed for the Albury-Wodonga Growth Centre but whether these will now eventuate following curtailment of the "growth hormone" remains to be seen.

Despite this obvious interest in transition and the many assumptions made about transition problems, the Curriculum & Research Branch survey referred to above is the only study located which focuses upon the primary-secondary interface in Victoria. Furthermore little research has been reported elsewhere.

One well-known investigation is that of Nisbet & Entwistle (1969) who looked at the attitudes and achievement of students transferring from primary to secondary schools in Aberdeen. Their study used a series of attitude and achievement tests that were completed by teachers and pupils. The findings indicated that the transfer affected different children in different ways. Pupils who came from educationally supportive homes found the transfer less traumatic than pupils from less academically oriented families. The Scottish pupils were also asked to write an essay about their impressions of the first day in the secondary school. Many of the essays described disturbing and frightening experiences. Six weeks later the pupils were asked to write another essay giving their impressions of secondary school. Comparisons between the two sets of essays suggested that after six weeks about eighty per cent of the children had adjusted to the new school and methods of teaching. It was noted, however, that some hostile attitudes still existed.

This paper describes a preliminary study envisaged as the first stage of a project designed to provide information on which to base the following judgments:

1. What undesirable effects, if any, are commonly associated with the transfer of children in Victoria from primary school to secondary school?
2. What is the incidence, magnitude and duration of any such undesirable effects?
3. How effective are existing programs and procedures aimed at reducing problems associated with transition from primary to secondary school?

The preliminary study is concerned with developing a satisfactory methodology by attempting to answer these questions (particularly the first two) in a limited context, with a view to a more representative study subsequently being carried out in a range of Victorian communities. Ultimately it is hoped to identify the characteristics of effective transition procedures.

## METHODOLOGY

An eastern suburban technical school (School X) and a neighbouring high school (School Y) were selected as the secondary schools for this pilot study. Four of their feeder primary schools (Schools A-D) were also involved in the study. Data relating to these schools are shown in Table 1.

The schools were chosen for the following reasons:

(a) All schools were fairly close to the researchers' places of work.
(b) Good contacts had previously been built up between the schools and the team members.
(c) Three of the primary schools sent pupils to both of the secondary schools.
(d) As far as could be ascertained the primary schools were representative of the total group of feeder schools.

(e) Both secondary schools drew from a wide range of socio-economic areas, were approximately the same size and were co-educational.

(f) School X had established a discrete transition school designed to ease the pupils into the new school situation, whereas School Y had only relatively minor provisions for reducing transition difficulties.

A stratified random sample of 44 pupils was selected as the group for detailed study (Table 1).

TABLE 1: Numbers in Grade 6 and Grade 7

| | | Secondary School | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | School X | | | | School Y | | | |
| Primary School | No. in Grade 7 / No. in Grade 6 | 183 | | | | 111 | | | |
| | | No. interviewed in Grade 6 | | No. interviewed in Grade 7 | | No. interviewed in Grade 6 | | No. interviewed in Grade 7 | |
| | | Boys | Girls | Boys | Girls | Boys | Girls | Boys | Girls |
| Primary A | 68 | 1 | 3 | 1 | 3 | 3 | 4 | 2 | 4 |
| Primary B | 84 | 4 | 4 | 4 | 4 | - | | - | |
| Primary C | 85 | 3 | 5 | 3 | 5 | 3 | 4 | 3 | 4 |
| Primary D | 31 | - | 2 | - | 2 | 4 | 4 | 2 | 4 |
| Total | 268 | 8 | 14 | 8 | 14 | 10 | 12 | 7 | 12 |

The data collection procedures were as follows:

(i) Pupil interviews

The children in the sample group were interviewed about two weeks before the end of their Grade 6 year and again mid-way through Term 1 of their Grade 7 year. Follow-up interviews are planned near the end of the Grade 7 year. Interviews were conducted by the investigators, with one assistant.

For the Grade 6 interviews an open ended schedule was devised which guided the interview whilst allowing unanticipated comments to be probed. This flexibility was seen to be desirable as the design of a written type questionnaire for use with larger samples was being considered.

A trial version of the interview schedule was tried at a nearby primary school and as a result the form was amended slightly. The interview sought to obtain information relating to:

(a) the home environment

(b) the pupil's perception of his class in particular and the primary school in general

(c) the pupil's expectations of his chosen secondary school.

A second interview schedule was subsequently designed for the follow up Grade 7 interviews. This schedule sought information corresponding to (b) above, but in the secondary school context, together with the pupil's perception of the transition in relation to his expectations.

(ii) Attitude Survey

At the times of interviewing an attitude survey was conducted of the whole Grade 6 or Grade 7 population of the school. The instrument selected was the School Sentiment Index, taken from the Instructional Objectives Exchange Publication

"Attitude to School" (1972). This instrument encompasses several of the attitudes in the interview schedule and has the advantages of being couched in simple language, of having been trialled and validated and of having been designed to be administered orally. Minor modifications, essentially to idiom, were made to the instrument. These were not considered likely to have any effect on its validity for ascertaining pupils' perceptions of:

- certain aspects of their teachers' behaviour
- their feelings towards learning related activities
- the social structure and climate of their school
- peer relations in their school context
- their school generally.

For each of the 81 items (read out by the tester to the whole grade), students were asked to indicate if an item was 'true' or 'untrue' to them and to indicate their answer on an answer sheet.

Three features of administration should be noted:

(a) The tester was a member of the project team. It was hoped that this would minimise "false" responses to questions which might occur if the test was given by the class teacher. The class teacher was in fact asked to remain out of the room.

(b) Each of the 81 statements was read to the class by the tester. It was hoped that this would overcome the difficulty faced by poor readers in answering the test. This also meant that more statements could be included than if the students had to read each statement.

(c) Students responded anonymously except that the interview sample were asked to put a cross in one corner of the answer sheet. A symbol to designate sex was also added.

Seven subscales are contained in the School Sentiment index and separate scores were computed for each student.

The resulting scores represent a measure of the pupil's perception (positive or negative) about each of the following aspects of the school.

1. The Teacher's Mode of Instruction (TMI) - adequacy, appropriateness, fairness of instruction and grading. Sample statement: "My teacher expects too much when he corrects my work."

2. The Teacher's Authority and Control (TAC) - fairness, effectiveness, manner, allowance for pupil participation. Sample statement: "My teacher allows us to make many decisions together in class."

3. The Teacher's Interpersonal Relationships with Pupils (TIRP) - consideration, friendliness and concern. Sample statement: "My teacher listens to what I say."

4. The Student's Attitude Toward Learning (Lg.) - interest and/or involvement in learning, attitude toward homework, different activities, independent learning. Sample statement: "I like to do some of my work at home."

5. Pupil's Perception of the Social Climate and Structure of the School (SSC) - organisation, bureaucracy, tradition. Sample statement: "The adults at my school are friendly."

6. Peer Relations in the School Context (Peer) - friendship patterns, equality, acceptance, stratification. Sample statement: "Other students often get me into trouble at school."

7. General Notion of School (Gen.) - perception of the holding power of the school. Sample statement: "Most school days seem as though they will never end."

(iii) Teacher Reports

Information was collected from teachers in two ways:

(a) Soon after interviewing each Grade 6 student, the team member concerned discussed the student with his teacher. Teachers were asked to comment on likely transition problems.

(b) Teachers were asked to complete an information schedule about each pupil interviewed. Information was sought about the pupil's academic status, his aptitude and ability, and factors likely to influence his smooth transfer from Grade 6 to Grade 7. When students were interviewed in Grade 7, their teachers completed corresponding information schedules but no informal interview was held.

(iv) Parent Reports

Two methods of obtaining parent data were tried. This information was collected a short time after the Grade 7 interviews.

(a) One team member telephoned a random sample of parents of those children interviewed. Parents were asked to recall how their child had transferred any problems he had, what the child had said about his new school and what the parents thought of the school.

(b) A questionnaire was devised, based on the results of the telephone interviews. These questionnaires were mailed to those parents who were not telephoned.

Figure 1 summarises the data collection procedures.

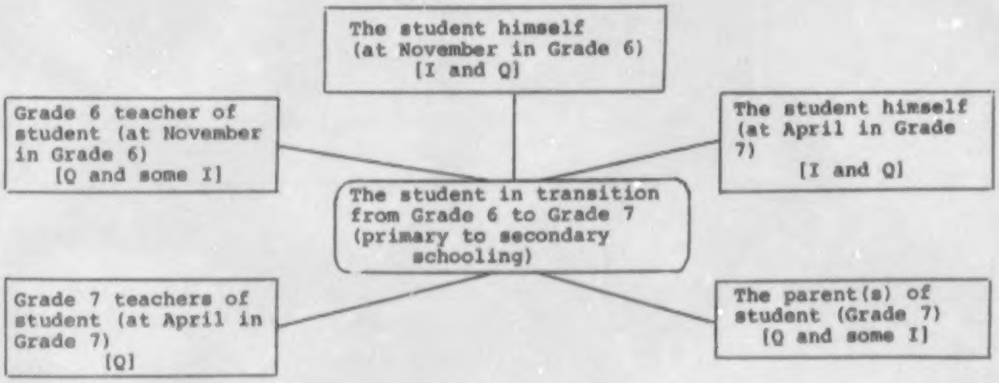

The student himself (at November in Grade 6) [I and Q]
Grade 6 teacher of student (at November in Grade 6) [Q and some I]
The student himself (at April in Grade 7) [I and Q]
The student in transition from Grade 6 to Grade 7 (primary to secondary schooling)
Grade 7 teachers of student (at April in Grade 7) [Q]
The parent(s) of student (Grade 7) [Q and some I]

Figure 1: Sources of information and methods of data collecting employed

## RESULTS AND INTERPRETATION

For each individual the information obtained from all sources provides a unique picture and one could present the findings as a series of individual case studies. This, of course, is not feasible in this paper. What will be reported instead are those findings relating to trends which seem to have emerged and which highlight areas of possible further research.

### (i) Results from Student and Parent Interviews

The Grade 6 interviews left no doubt that students were positively anticipating proceeding to secondary school. The major reason for looking forward to the transfer was that it would be a 'new experience'. Several students indicated that after being at the same school for seven years they were looking forward to a change of school and to having more teachers, more space and enticing new subjects. For many the prospect of being able to take woodwork, sheetmetal, textiles and cooking was a major attraction.

TABLE 2: Anticipated and Actual 'Likes' Associated with the Grade 7 School

| QUESTION | | Grade 7 School | |
|---|---|---|---|
| | | School X | School Y |
| | | (expressed as percentages) | |
| 1. At November in Grade 6 | | | |
| (a) Are you looking forward to going to secondary school? | yes | 91 | 73 |
| | no | 5 | 5 |
| | uncertain | 5 | 23 |
| (b) What are you looking forward to at your secondary school? | sheet metal )<br>woodwork )<br>textiles )<br>cooking ) | 45 | 19 |
| | new experience | 45 | 27 |
| | new friends | 14 | 5 |
| 2. At April in Grade 7 | | | |
| (a) Are you glad you came to this school? | yes | 73 | 84 |
| | no | 14 | 11 |
| | undecided | 14 | 5 |
| (b) What do you like most about your secondary school? | woodwork | 33 | 16 |
| | sheetmetal | 23 | |
| | textiles | 33 | |
| | art & craft | 36 | |
| | home economics | 27 | |
| | P.E. | 18 | |
| | drama/English | 14 | 16 |
| | sport | 14 | 21 |
| | activities during lunch break | 14 | |
| | workshop | | 16 |
| | new friends | | 16 |
| | library | | 10 |

It can be seen from Table 2 that the majority of students were glad that they had come to their particular secondary school, with the manual subjects and sport being strongly linked to 'most liking' the school. Parents' comments on their children's report of what was liked at school support this. When in Grade 6, the most important 'likes' about primary school according to the sample were also sport and art and craft (50% and 27% respectively of the total sample).

When asked in the Grade 6 interviews what they were not looking forward to at secondary school, very few students had any response. Each of the following was mentioned only once .... different teachers, uniform, swapping rooms, difficult work, being youngest in the school .... while school size and loss of friends were nominated only twice. When expressly asked 'Do you have any worries about going to secondary school?' approximately half indicated 'no worries'. Worries reported by others included

- probably picked on by older students (15%)
- teachers might not be so helpful (12%)
- have to make new friends (7%).

It therefore appears that for students the expectation of a 'new experience', together with opportunities to participate in manual subjects outweighed their anticipated dislikes and worries. This impression was strengthened on the one hand by the students in Grade 7 reporting very few dislikes .... maths, science, some teachers and too far to walk .... being the major ones, but each reported by only a couple of students, and on the other hand by parents indicating that of sixteen given words the ones which would have most applied to their children in the first few weeks of school were .... uncertain (56%), excited (39%), calm (39%), enthusiastic (39%). Parents also reported that sport (100%), manual subjects (83%), having different teachers (67%) and constant room changes (44%), were their children's 'likes' of the secondary school, with few dislikes reported, the most frequent being teasing from older children (39%).

TABLE 3: Perception of Work in the Grade 7 School and Anticipated Teacher Differences

| QUESTION | | Grade 7 School | |
|---|---|---|---|
| | | School X | School Y |
| | | (expressed as percentages) | |
| 1. At November in Grade 6 | | | |
| (a) What do you think secondary school will be like? | harder | 32 | 27 |
| | homework | 23 | 14 |
| | different teachers | 9 | 23 |
| | youngest in school | 14 | 18 |
| | swapping rooms | - | 14 |
| (b) Do you think the teachers will be any different than your Grade 6 teachers? | stricter | 32 | 23 |
| | make you work harder | 9 | 27 |
| | won't know you as well | 9 | 9 |
| | no real difference | 50 | 45 |

TABLE 3 continues

| | | School X | School Y |
|---|---|---|---|
| 2. At April in Grade 7 | | | |
| (a) Do you find the work harder in Grade 7 than in Grade 6? | yes | 35 | 74 |
| | no | 20 | 5 |
| | same | 45 | 21 |
| (b) Do you wish the work was harder? | yes | 74 | 29 |
| | no | 26 | 71 |
| (c) Are you interested in your work? | yes | 85 | 79 |
| | no | - | - |
| | undecided | 15 | 22 |
| (d) Are you more interested in your work in Grade 7 than when in Grade 6? | yes | 80 | 63 |
| | no | 5 | 11 |
| | same | 15 | 26 |

In Table 3 some interesting trends are evident, particularly between the two secondary schools. For students in Grade 6 the most common expectation they had of secondary school was that the work would be harder. For School X some Grade 7 students (35%) indicated the work was harder but the majority (74%) indicated that they wished it was harder. These results are reversed in the case of School Y.

Parents perceive that in the secondary school, as compared with the primary school:

| | |
|---|---|
| . the expected standard of pupil behaviour is | - higher (89% X, 56% Y)<br>- about the same (11% X, 44% Y) |
| . the expected standard of written work is | - higher (77% X, 67% Y)<br>- about the same (23% X, 33% Y) |
| . the amount of concern shown for individuals seems to be | - greater (33% X, 44% Y)<br>- about the same (56% X, 33% Y)<br>- less (11% X, 23% Y) |

Students on the whole viewed the Grade 7 teachers to be not that dissimilar to those in Grade 6. Comments which were forwarded for School X included ... treating us more like adults, understanding you better, help you a lot, seem to like you, different, explain things more ...., while for School Y they included ... stricter, don't know you, not as friendly, better, get fed up more easily, explain things better .... This supports the significantly lower 'Teacher's Interpersonal Relationship with Pupils' (TIRP) score in School Y compared with School X (see Table 4).

Students perceived the classroom environment in Grade 7 to be generally similar to what they had experienced in Grade 6 although greater restriction on movement within the classroom was perceived and detention was now a form of punishment.

### (ii) General Impressions Gained from Student and Parent Interviews

The general picture emerges that for this sample of students, transition from Grade 6 to Grade 7 had been viewed with positive anticipation and not a great deal of concern. It was expected that the work would be harder in Grade 7 and the teachers would be stricter, although not that different. Students in School Y generally found this to be the case, but not so for students in School X, where there was a positive expression of wishing the work would be harder.

It is relevant to note here that

- approximately half of the sample had brothers or sisters already attending the school
- only one student reported no friends going to the secondary school
- almost 80% of the students had not wanted to go to a different secondary school.

These factors tend to provide support for the student in the secondary school.

### (iii) Results from the Attitude Survey of the Sample

Results for the interview sample on each of the factors of the School Sentiment Index are shown in Table 4.

TABLE 4: Factor Scores for the Sample at Both Primary and Secondary School

| Factor | Possible Score Range | Primary Schools (combined) | School X | School Y |
|---|---|---|---|---|
| TMI | 12-24 | 21.51 | 21.33 | 20.63 |
| TAC | 12-24 | 19.19 | 18.44 | 18.79 |
| TIRP | 11-22 | 18.40 | 18.17 | 16.92 |
| Lg. | 10-20 | 15.78 | 15.95 | 16.17 |
| SSC | 12-24 | 20.51 | 19.67 | 19.13 |
| Peer | 12-24 | 20.49 | 19.83 | 21.58 |
| Gen. | 12-24 | 18.08 | 18.56 | 17.50 |

The closer the actual score to the maximum score, the more positive is the attitude towards that factor expressed by the students.

Generally scores on the factors are not as high at the secondary schools as at primary school. It is of interest to note the consistency between the interview data and the attitude survey data. For example:

- the relatively low factor score on Teacher's Interpersonal Relationships with Pupils (TIRP) by the students at school Y and the relatively higher score on PEER by these students, the PEER score being higher than for the sample combined when at primary school (cf. student comments on teachers, presented earlier)
- the more positive attitude towards learning (Lg.) at both secondary schools compared to the primary school (cf. Table 3)
- the more positive attitude towards school in general (Gen.) indicated by students at school X compared with the primary school and school Y (cf. Table 3).

### (iv) Results from Teacher Questionnaire Returns and Interviews

From the questionnaire returns and interviews of the Grade 6 teachers, those characteristics of the individual student which were expected to have a negative influence on transition primarily related to social characteristics such as being shy, sensitive, easily distracted, unable to get on with peers, lazy. These were viewed as being more important than weakness in maths or English.

Grade 7 teachers viewed social characteristics similar to those above as having a negative influence, including also difficulty in interacting in a group. In addition, they commented on lack of ability in reading and/or English. However it was a deficiency in what could be described as 'satisfactory study habits' which was most often mentioned. Statements such as ... unable to organise adequately, short attention span, keen to start but very slow, not concerned about presentation of work ... were often made.

The data concerning each individual child in the sample were examined in totality. None of the children appeared to be experiencing major difficulties in transferring. Those concerns that were mentioned by individual pupils were almost entirely minor and short-lived. Information from parents was consistent with that obtained from the children themselves.

What did become apparent was that "having problems" at transition from primary to secondary school, in the words of the popular song, "means different things to different people". The question then emerges: on whose perception will problems of transition be considered to be present? This issue is commented upon in the discussion later in the paper.

### (v) Attitude survey results for all students in Grade 6 at the four primary schools and all students in Grade 7 at the two secondary schools

An analysis of variance was carried out on the data for the four primary schools. The ANOVA was significant for all factors (at <0.01 level), and tests for pairwise comparison showed in the main only one significant difference per factor. The average factor scores for each primary school are shown in Table 5, together with the combined averages scores. Included in the Table are the average factor scores for each secondary school.

TABLE 5: Average Factor Scores for Each School in the Study

| School | When collected | N | TMI | TAC | TIRP | Lg. | SSC | Peer | Gen. |
|---|---|---|---|---|---|---|---|---|---|
| | | | (max24) | (max24) | (max22) | (max20) | (max24) | (max24) | (max24) |
| Primary A ) | | 68 | 22.22 | 19.62 | 19.81 | 16.72 | 20.25 | 20.10 | 19.99 |
| Primary B ) | Grade 6, | 84 | 20.56 | 18.40 | 17.38 | 15.38 | 19.24 | 19.39 | 17.82 |
| Primary C ) | November | 85 | 21.97 | 19.49 | 18.20 | 16.82 | 20.92 | 19.94 | 19.35 |
| Primary D ) | 1976 | 31 | 21.58 | 19.52 | 18.22 | 16.10 | 20.33 | 20.77 | 16.68 |
| Combined Primary | | 268 | 21.54 | 19.18 | 18.35 | 16.26 | 20.07 | 19.91 | 18.72 |
| Secondary X) | Grade 7, | 183 | 20.76 | 17.78 | 17.54 | 15.68 | 19.42 | 20.44 | 17.73 |
| Secondary Y) | April 1977 | 111 | 21.16 | 18.50 | 17.47 | 16.13 | 19.13 | 20.59 | 17.25 |

Differences between the primary schools are evident on the factor series with Primary A in the main obtaining the highest scores and Primary B the lowest scores. The significant differences between the combined primary schools and the two secondary schools are indicated in Figure 2.

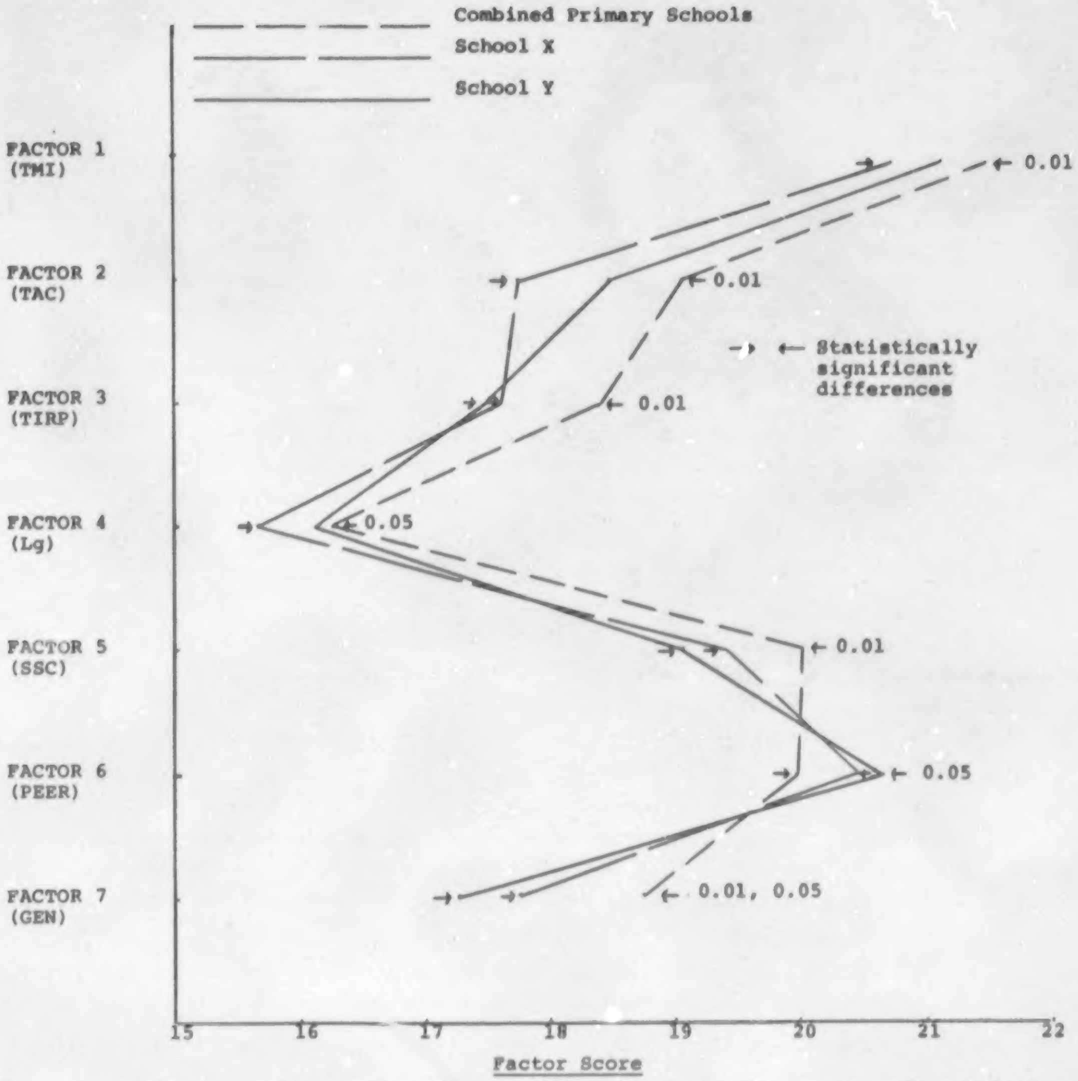


Figure 2: Factor scores for combined primary and the two secondary schools

A 't' test was used to compare the factor scores of the pupils in the interview sample at each secondary school with those of the remainder of the Grade 7 population of that school. With the exception of the factor SSC, there were no significant differences. It was therefore assumed that the student sample in each of the two secondary schools was representative of the total school Grade 7 population with respect to attitude to school.

## DISCUSSION

This preliminary study was carried out primarily as an attempt to clarify appropriate form and dimensions for a more comprehensive future investigation, and to develop suitable information-gathering procedures. To these ends it was important that the general approach should be open and responsive and that a range of data-collecting methods should be tried.

It was considered essential to attempt to find ways of exploring individual children's own thoughts, feelings and concerns as they transfer. It is after all, the individual's perception of a situation which will result in concern or anxiety being experienced by that individual. On the other hand, a "problem" may exist which is not a self-perceived concern. Hence it was considered important to gather information from others close to the child. Teachers and parents were considered to be the most suitable sources. Peers and/or siblings might be contemplated, but there could be dangers with using these.

(i) Comments on the various techniques used

(a) Pupil Interviews: The major reasons for using interviews were to allow for more personal and open-ended responses and to avoid reading and writing difficulties. The interview team consisted of two males and two females. Each member adopted an empathetic approach and in questioning pupils, initially sought comments of a positive nature before exploring negative feelings. All members noted that quite a few students had some difficulty in verbalising their thoughts. The team found no difficulty so far as pupil openness and friendliness were concerned. Most pupils were perceived to be relaxed and quite happy to talk about their school. Each individual pupil was interviewed by the same team member on both occasions.

Some pupils were found to be reluctant to initiate comments. Others were inclined to "defend" their school, which may have prejudiced some of their comments.

It is suggested that less direct interview techniques may be useful with this age range. One possible method would be to show photographs or drawings of a pupil in a variety of situations around the school and ask the question "What do you think he/she might be feeling?" It is proposed to experiment with this technique.

(b) Pupil Attitude Survey: As stated earlier the School Sentiment Index was used as an additional means of testing several of the attitudes towards school that were included in the interview. The pupils gave no indication of having any problems associated with either understanding the orally presented statements or recording their responses.

The administrative procedures adopted were perceived by the team to be very satisfactory and to have the following advantages:

(i) The opinions of many pupils about many statements could be obtained in a short time.

(ii) The teachers did not object to leaving their classes for 20 minutes but we suspect more resistance would have occurred if longer times were requested.

(iii) The oral presentation and Yes/No answers overcame many of the problems related to reading ability and written expression.

(iv) The children were probably more honest in their answers due to the absence of their teacher.

It is felt that in future studies it would be desirable to be able to individually identify the interviewed sample by having them write their names on their answer sheets. It is unlikely that students would mind doing this: some did so anyway, presumably from force of habit. Data obtained with this instrument are seen as useful in that they provide one means of checking representativeness of a smaller sample against the whole school population in terms of attitude to school. In addition, the results would provide one useful set of criteria for gauging the effectiveness of a school's transition procedures.

(c) Teacher information: Teacher comments about pupils were sought either by direct interviews with teachers or by a written questionnaire. Teacher-based information is valuable for three reasons:

- (i) Students considered likely to experience transition problems can be identified in Grade 6.
- (ii) Students perceived as having difficulties in Grade 7 can be identified.
- (iii) The student's perception of his ease of transition can be compared with that of another observer.

It was natural that the primary teachers knew more about their pupils in December than did the secondary teachers in April. However the secondary staff were able to provide useful information about many of the sample.

The direct interview technique was found to be more effective and avoided the negative attitudes related to form filling. If one member of the team spoke to the teacher whilst another team member administered the attitude survey to the class the teacher would be subjected to the least amount of inconvenience.

(d) Parent information: Information about the transition of the sample students was sought from their parents. Half the families were telephoned and the other half were sent a questionnaire based on the data from the phone calls. The telephoning of parents, mostly mothers, was time consuming but more valuable information seemed to come from these calls. Where families have no telephone, substituting home interview visits would be expensive of time.

(ii) Implications for Wider Study

Overall the information supplied by parents, teachers and children matched fairly closely. It seemed that no child in our sample experienced any great transition problems. The hypothesis can be advanced that only a small percentage of Grade 6 pupils do experience difficulty. This suggests that for exploring the nature (rather than the incidence) of transition problems the Grade 6 teachers should be asked to predict those pupils expected to have difficulties rather than interviewing random samples. These students could then form the sample group to be interviewed.

Less difference was found between the classroom practices of Grade 6 and 7 teachers than was anticipated. This may have been a major contributing factor to the smooth transition observed. It would seem likely that the differences between Grade 6 and Grade 7 classrooms which may have aggravated transition difficulties in the past are tending to disappear. Movement from a one-teacher situation to a multi-teacher situation is generally no longer an accurate description of the current transfer to secondary school, either. Grade 6 students are likely to have up to five or six different teachers as a result of specialisation in art and craft, physical education, music, library and sometimes other fields.

These comments must be tempered by the complicating factor of the open-space primary schools currently being built in Victoria. While practices within these vary, the transfer from an open area primary school to a conventional secondary school context is likely to be a more drastic change than between the more conventional schools involved in the preliminary study. Consequently an extended

370

study would need to include this situation and the reverse one where a conventional primary school feeds into an innovative secondary school.

While the schools studied did draw from a rather diverse community socio-economically, it is recognized that a quite different situation might prevail in, for example say, an inner-suburban area, or a rural area. It is clear that future studies would need to include a representative range of socio-economic areas in their sample as our result may be atypical.

The team's contact with independent schools suggests that there may be a considerable difference between classroom practices of Grade 6 in state schools and Grade 7 in many independent schools. Additionally, the greater travelling often involved, coupled with the reduced likelihood of peer-group associates accompanying the individual to the new school, would seem likely to aggravate transition difficulties. Thus, in a more extensive study, independent schools could be included.

Finally transition would need to be studied in schools where there is a range of degrees of provision designed specifically to ease transition. This should enable some conclusions to be drawn about the extent of provision appropriate in various situations.

The results presented in this paper are both preliminary and representative of a very small sample of schools. It may be that further data reveals a somewhat different picture. However, if it was to be established that most students transfer into the secondary school without experiencing major problems, this would still be valuable information. Schools which have developed extensive transition programs may, in this case, be better advised to invest equivalent resources in other ways.

## Bibliography

Dutch, R.D. and McCall, J. Transition to secondary - an experiment in a Scottish Comprehensive School. British Journal of Educational Psychology, 1974, 44, 282-289.

Educational Facilities Laboratories. Middle School. A report of two conferences held at Mt. Kisco, N.Y., September 29-30, 1961. New York: The Laboratories, 1962.

Instructional Objectives Exchange. Attitude Toward School, Grades K-12. Los Angeles, 1972.

New South Wales, Education Department. Report of the Committee Appointed to Survey Secondary Education in New South Wales (H.S. Wyndham, Chairman). Sydney: The Department, October 1957.

Nisbet, J.D. and Entwistle, N.J. The Transition to Secondary Education. London: University of London Press, 1969.

Scottish Education Department. Secondary Education - A Report of the Advisory Council on Education in Scotland (W. Hamilton Fyfe, Chairman). Edinburgh: Her Majesty's Stationery Office, 1947.

Telfer, P. Jr. Middle Schools. In L.C. Deighton (Ed.), The Encyclopedia of Education, Vol. 6. New York: Macmillan Co. and The Free Press, 1971.

Victoria, Education Department. Primary-Secondary Transition (Research Report). Melbourne: September 1966.

Victoria, Education Department. Report of the Committee on State Education in Victoria (Alan H. Ramsay, Chairman). Melbourne: The Department, February 1960.

Western Australia, Education Department. Secondary Education in Western Australia (H.W. Dettman, Chairman). Report of the Committee on Secondary Education. Perth: The Department, February 1969.

## SUNDAY 13 NOVEMBER

## 9.45 - 10.30am

| PAPERS | Rooms |
|---|---|
| D. Williams (Canberra C.A.E.)<br>The Design, Development and Evaluation of a Curriculum: The Aboriginal Australian in N.E. Arnhem Land | Gateway Room |
| OMITTED WORKSHOP ON TEACHER EDUCATION | Ladies Drawing Room |
| J. Braithwaite (Macquarie University)<br>Evaluation of Early Childhood Programs | Seminar Room B |
| R. Docking (Murdoch University)<br>Anxiety, Achievement and Antipathy | Seminar Room D |
| B. Fraser (Macquarie University)<br>OMITTED Research into the Evaluation of ASEP | Meeting Room |

# THE DESIGN, DEVELOPMENT AND EVALUATION
# OF A CURRICULUM:
# THE ABORIGINAL AUSTRALIAN IN NORTHEASTERN ARNHEM LAND

by

Don Williams*
(Canberra C.A.E.)

## RATIONALE AND AIMS

In Australia at present there is a tremendous amount of interest in Aboriginal studies in primary, secondary and tertiary institutions. During the last few years protagonists of the Aboriginal cause have endeavoured to draw attention to racially biased textbooks and replace them with up-to-date, accurate texts. It is significant to note that the newly established Curriculum Development Centre has made Aboriginal Education one of its areas for immediate action. There is a clear need for first-class materials about Aborigines. As educationists and teachers perceive this need they are responding by producing resource kits of varying quality and scope. This is a desirable development, particularly when such resource kits are developed by teachers to study the Aborigines in their locality. In passing, it is appropriate to note that good curriculum materials are required about all cultural groups in the Australian society.

The Aboriginal Australian in Northeastern Arnhem Land is planned to be something more than a box of materials about a group of Aborigines. To start with, it is an extremely detailed set of multi-media materials containing, for example, about forty separate booklets. It is detailed and comprehensive because it does not wish to create the idea in the minds of people that a few weeks' work will be adequate to gain an understanding of Aboriginal life. Furthermore, nothing short of an in-depth study of people as people is likely to produce significant cognitive and attitudinal outcomes. Platitudes about Aborigines' ability to hunt and throw spears coupled with emotionally charged comments about racism in modern Australia are grossly inadequate. They are patronizing to say the least but their long-term effect could be more insidious. The so-called attitude change they purport to produce is likely to be superficial because it lacks a cognitive underpinning. The Aboriginal Australian in Northeastern Arnhem Land deals with more than just the artifacts of culture; it is a substantial course in anthropology for I2-I4 year old children. A challenge to evaluators of the curriculum in future years will be to examine this assumption upon which the curriculum is predicated.

If the curriculum were simply a comprehensive and in-depth study of the Aborigines in Northeastern Arnhem Land and nothing more than this, it would be deficient when used in classrooms throughout Australia. It needs to be a mirror in which pupils see themselves, their society and their culture. When studying Aboriginal religion, children need to reflect on their own religious experience. When learning how Aborigines dispose of a dead body, children need to reflect on the way their community handles death. "What do other peoples throughout the world do?" is a logical extension of this line of enquiry. It is not simply a matter of "seeing ourselves as other see us" but of "seeing ourselves in others". Because the neo-traditional Aboriginal life in Northeastern Arnhem Land is so different from the dominant White Australian lifestyle it should act as a fairly bright mirror in which White Australian children can see themselves.

Can a detailed study of a group of Aborigines, incorporating in the process a study of one's own way of life, be justified at the end of primary schooling or at the beginning of secondary schooling? The curriculum is being developed on the assumption that this question will be answered in the affirmative by a large number of parents and teachers. It is designed to synthesize the knowledge, skills and attitudes which children have acquired in the social sciences and humanities during the course of their primary education. If teachers and parents feel that children should know about supply and demand in economics, power and decision-making in politics, ecosystems and environmental impact studies in ecology, family structure and kinship in sociology, phonemes and dynamic equivalents in language, etc., this curriculum should appeal to them. Such a claim may seem to suggest that the universal panacea for all educational ills has at long last been found. The curriculum is no panacea but it is a well-planned case study. It pulls together key concepts from a variety of disciplines and involves children in modes of enquiry used in the disciplines. If the curriculum were a well-designed case study about a group of people other than Aborigines, there would be sound educational reasons for incorporating it in a school curriculum to synthesize learnings in the social sciences and humanities.

---

* The Curriculum is being developed jointly by Don Williams and Alan Fidock, who is the coordinator of Aboriginal projects at the Curriculum Development Centre. The authors are working in close consultation with the town councils in North Eastern Arnhem Land and, through the Curriculum Development Centre, which is funding the project, will also work with representatives of the recently established National Aboriginal Education Committee.

In the U.S.A., Canada, many European countries and Australia there is a great deal of discussion about the type of education required for life in a culturally pluralistic society. It is hoped that the Aboriginal Australian in Northeastern Arnhem Land will make a valuable contribution in preparing Australian children for life in a pluralistic society because it has been developed in the belief that a liberal education for each child should include a study of a culture or sub-culture different from his or her own. Australians need to get to know each other in all their cultural complexities. They need to be equipped with the knowledge and skills to appreciate different life-styles.

To sum up, the curriculum is designed for people other than the Aborigines in Northeastern Arnhem Land with the following aims in mind:

1. To make a detailed and comprehensive study of Aborigines in Northeastern Arnhem Land and hence improve the knowledge people have about Aborigines and the attitude they hold towards them.

2. To enable pupils to examine themselves, their society and their culture as they study the Aborigines in Northeastern Arnhem Land.

3. To synthesize the learnings which children have acquired in the social sciences and humanities by means of a case study.

4. To prepare children for life in a culturally pluralistic society.

## DESIGN OF THE CURRICULUM

In this brief paper it will be impossible to discuss all the principles which went into the design of the curriculum. A few of the basic planning concepts will be mentioned and an over-view of the design presented.

When selecting content, two basic questions were asked: what do Aborigines in Northeastern Arnhem Land want other people to know about them and what do White Australian children want to know about Aborigines? Answers to these questions were not sought in rigorous quantitative manner. In all the discussions with Aborigines and White Australian children they have been guiding principles. By writing booklets, several Aborigines have made direct contributions to the curriculum which will be subject to scrutiny and approval by the town councils in Arnhem Land.

Because most teachers and their pupils will know little about the Aborigines in Northeastern Arnhem Land and will have limited access to resources about them, the curriculum had to be planned as a self-contained unit suitable for use in all kinds of classrooms. Small units of work, individually packaged and grouped into modules, were therefore prepared to permit maximum versatility in the use of the materials.

The curriculum caters adequately for pupils of differing abilities and attainments. Poor readers, for example, will have books they should be able to handle and will be able to supplement their reading with audio/visual aids. The high achievers have not been neglected. A complete module on kinship and social organization has been included to extend the performance of the most capable of the students.

Several possible designs for arranging the content were considered before a decision was taken to follow a fairly traditional approach. In the higher levels of the curriculum the content is organized around such disciplines as history, ecology, politics and economics. This traditional type of organization is still advocated by many curriculum theorists. It certainly makes good sense in the majority of Australian high schools. There are dangers in organizing content around social issues, for example, and presenting Aborigines as "problems" in modern Australia. While the overt structure is discipline-oriented, there are several covert substructures. For instance, racism, the problem of alcoholism and land rights are adequately dealt with as social issues across the various modules in the curriculum.

It is somewhat trite to claim that the curriculum provides for a variety of teaching and learning styles. No sub-script, such as "an inductive approach", can be appended to the title as there is no distinctive mode of teaching/learning pervading the curriculum. This does not imply that the generation of materials, which contain suggested learning activities, has been undertaken without regard to modes of teaching and learning. On the contrary, as the curriculum has been organized around disciplines, an attempt has been made to include the major modes of enquiry characteristic of the disciplines. Activities in the booklets range from programmed instruction in basic kinship symbols through to a process-oriented enquiry into social issues using strategies developed by Stenhouse in the Humanities Curriculum.

The above comments should suffice to show why the following outline of the content was adopted.

### LEVEL I: SETTING THE SCENE (Orange Module)

In this short introductory module, which pupils are expected to complete as a group or class, a variety of media are used to arouse motivation and lay a foundation for subsequent studies. As Level 2 involves an analysis of activities in which Aborigines participate, it is important for each major activity to be featured in the introductory level. Pupils will then be in a position to pursue their interests on an individual basis. Titles in this module are as follows:

- Study Guide (book)
- A Week in the Life of John Djan'kawu (book)
- A Week in the Life of Lorraine Lambu (book)
- Arnhem Land Panorama (slides/cassette)
- Arnhem Land People (slide/cassette)
- Arnhem Land Pictures (wall pictures)

After the initial trialling and evaluation of the curriculum, it is hoped to produce a movie film to supplement the material in Level I.

### LEVEL 2: ANALYZING ACTIVITIES (Blue Module)

A small booklet is included in multiple copies (five) for each of the activities and, where appropriate, supporting audio/visual materials are included. During the development of the curriculum, a wealth of audio/visual material was collected to both supplement and complement the written materials, but the cost structure has placed severe restraints on a multi-media approach. Some of the 16 titles in this module are as follows:

- Children's Games
- Dhapi Initiation Ceremony
- School and Adult Education
- Fishing Industry
- Bark Painting
- Mining
- Agriculture
- Local Government
- Major Ceremonies
- Church

Suggested learning tasks which emanate from the studies of Aboriginal activities are included in the booklets. At this level these tasks are designed to explore and analyze Aboriginal activities in such a way that a need is established to proceed beyond analysis toward a synthesis of knowledge in the Level 3 materials. For example, when studying mining, pupils should generate a need to know more about clans. An analysis of bark painting will call for a knowledge of religion. It is hoped that pupil tasks, such as doing a bark painting, will be stimulating.

### LEVEL 3: SYNTHESIZING KNOWLEDGE (Gold Module)

When human beings have synthesized knowledge, they have usually found it convenient to organise it into disciplines or major domains. When solving problems, they have usually drawn on knowledge from several disciplines or domains, leaning heavily on one or two in particular. This module in the curriculum incorporates these two approaches to knowledge. Key concepts and inter-relationships between them are explored for each of the domains of knowledge listed below, and within each domain social problems are explored. For example, ecological concepts are established and used to explore environmental impact studies. The eight domains of knowledge are as follows:

- History
- Ecology
- Life Cycle
- Economics
- Politics
- Religion
- The Arts
- Kinship (An Introduction)

It is at this level that comparisons and contrasts are made between the Aboriginal way of life and the way of life of the pupils. For instance, pupils study the physical and psychological development of Aborigines in the booklet on "Life Cycle" and reflect on their own development through a series of learning tasks.

LEVEL 4: EXTENDING KNOWLEDGE (Green Module)

This is an extremely difficult module on kinship and social organisation which will extend the brightest of the pupils. The components are:

- Kinship Symbols
- Parents
- Grandparents
- Marriage
- Clans
- Moieties and Subsections
- Culture Contact

LANGUAGE (Purple Module)

Language is a key to understanding culture. An outstanding feature of the curriculum is the language course which may be studied concurrently with the materials in Levels 2 to 4. The components are:

- Learning About Language
- A Short Course in Gupapuyngu: Lessons I to IO
- A Short Course in Gupapuyngu: Lessons II to 20
- A Short Course in Gupapuyngu: Lessons 2I to 30
- A Short Course in Gupapuyngu: Lessons 3I to 40
- Literature from Northeastern Arnhem Land

TEACHERS' HANDBOOKS

Besides a general handbook for the whole curriculum, a separate handbook is available for the language module.

DEVELOPMENT AND EVALUATION

The remaining sections of this paper will describe how two modules of the curriculum (Level 4 and Language) were written and are still being developed and evaluated. As Taba et al (1971, p. 36 ) have shown, the curriculum developer usually plans from the highest levels of generalization down to basic facts and information which are the starting points for pupils. The modules in the curriculum are being completed in the following order:

I. Language
2. Level 4: Kinship and Social Organization
3. Level 3: Analyzing Activities
4. Levels 2 and I: Analyzing Activities and Setting the Scene

A DESCRIPTION OF THE LEVEL 4 MODULE

This module, which is an advanced study of kinship and social organisation, contains material which has traditionally been taught at tertiary institutions. Furthermore, there is by no means agreement among anthropologists about the nature and significance of fundamental aspects of the terminological and sociocentric kinship systems. In fact, the anthropology of this area is known in the literature as the "Murngin Controversy". In a monograph Professor Barnes, formerly Professor of Anthropology at the Australian National University, chose to explore aspects of this controversy under the title "Inquest on the Murngin". It is important for pupils to realise that this so-called simple primitive society has tested the minds of some of the world's leading anthropologists and to date seems to have defied them in the more esoteric areas.

A study of kinship is ideal for demonstrating to pupils how far modern technological society has swept people along. Aborigines, possibly more than any other people in the world, have maximized the significance of kinship. Kinship features in almost every aspect of life and is a basis for social structure and local organization. By way of contrast, modern White Australian society, along with other Western societies, has possibly minimized the significance of kinship more than any other society in the history of the world. The isolation of the nuclear family, the growing incidence of single parent families and the increased mobility due to modern transport and communications have reduced the significance of kinship as a basis for human interaction and relationships. The current emphasis on group life in churches throughout Australia, with jargon such as relating, sharing, affirming one another and the like, is indicative of the need which many White Australians face because the society can no longer base most of its human relationships on kinship. A study of kinship and social organisation in an Aboriginal community is therefore an effective way of teaching White Australian children about the basis of human relationships in their own society which contrasts so vividly with an Aboriginal society.

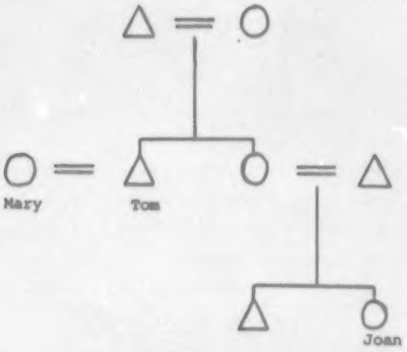
In the light of the above comments about Aboriginal kinship, it is obvious that this module should have features which are markedly different from the earlier modules. As the content calls for high levels of conceptualization, the curriculum materials are designed for children who are demonstrating the capacity to handle formal operations and fairly abstract thinking. Whereas the Level 3 booklet on kinship and social organization use stick figures to depict people, the booklets in this level use symbols and show how the abstract sociocentric kinship systems are generated. Kinship symbols are introduced in the first booklet which has a programmed learning format so that bright students may teach themselves.

The second booklet on "Parents" illustrates how the module attempts to let children see themselves in relation to other human beings. In most Aboriginal societies, mother's sisters are called "Mother" and any children they have are regarded as brothers and sisters. In other words, parallel cousins are classified as siblings which is the basis of classificatory kinship. Likewise, father's brothers are called "Father" and any children they have are regarded as siblings. The Aboriginal system is compared and contrasted with the dominant White Australian system and then children are invited to think of other possibilities for naming and classifying kin in the first ascending generation. They are shown that there are at least a couple of other main ways that people throughout the world use to refer to these kin.

The final booklet in the series of seven explores culture contact by examining what happens when a society that has maximized the significance of kinship meets another society that has minimized it. At the cultural level, the contrast is just as pronounced because a rich self-contained culture is in contact with a highly-developed diffused culture which people sample differentially. By using raw data from empirical research it is hoped that children will gain deep insights into the nature of contact, thereby better equipping them for life in a culturally pluralistic society. There are no easy answers to many of the problems associated with cultural pluralism, even if the problems are understood better. Appropriate teaching strategies are therefore required to enable pupils, as responsible well-informed citizens, to see how they may work with others towards the resolution of problems which are controversial.

From the above description of some of the features of the module, it is obvious that the development and evaluation of it present a challenge. As the materials are trialled in 1978, instruments will be used to determine whether pupils have mastered the difficult content. These will take the form of "Test Yourself Exercises" which appear in the pupils' booklets and an objective test which deals with higher order knowledge and processes. It is relatively easy to construct test items which explore higher order knowledge and processes as the following example shows:

Sample Item



Tom is in patrimoiety A and in matrimoiety Y.
Mary is in patrimoiety B and in matrimoiety Z.
Joan will be in:

☐ Patrimoiety A and matrimoiety Y
☐ Patrimoiety A and matrimoiety Z
☐ Patrimoiety B and matrimoiety Y
☐ Patrimoiety B and matrimoiety Z

It takes time to work out which is the correct answer. Preliminary work with a couple of bright students when the booklets were being written suggests that they are capable of handling this level of abstract thinking.

A far more difficult task will be an appraisal of how far the content of the module has enabled children to see themselves in relation to Aborigines and others. The most difficult task of all will be to determine how far the module has produced deep insights into the nature of cultural pluralism. Techniques for evaluating the effectiveness of the curriculum in these areas are currently being devised. In contrast to tests of knowledge about the content, they are open ended. Teachers' and pupils' perceptions of their own insights will be one line of enquiry while the opinions of a small group of professional evaluators who interact with the children at the conclusion of the module will be another. During the initial stages of the project an advisory panel consisting of Professor R. M. Berndt, Dr. D. Dufty and Miss M. Valadian examined some of the materials and provided useful guidance for the development of the curriculum. It is hoped that this panel will assist with a further evaluation of the curriculum when it is trialled during 1978.

A DESCRIPTION OF THE LANGUAGE MODULE

This is a foreign language course with a difference! Unlike foreign language courses in French, German, Italian and other well-known languages, it is a course which deals with a language:

- that is understood by about 3,000 people and spoken by under 1,000 people,
- that has only recently acquired a written form and as a result has a severely restricted written literature,
- that is unknown to the overwhelming majority of teachers and pupils who will study the language as part of the curriculum, and
- that is not generally accepted as a language worthy of study in primary and secondary schools.

Because the language under study possesses these characteristics, the course differs from other well-established foreign language courses. To start with, it is severely restricted in scope and does not set out to achieve high levels of mastery in the language. The aim is not to teach effective communication in the language but to do enough communicating in the language to discover how languages communicate. The emphasis is on learning about language through the partial learning of a language.

The starting point for the design of most foreign language courses is a concern to teach a language to such a level that it may be used effectively for general communication. The need to study aspects of culture in such courses is clearly recognized, but the primary concern is with language acquisition. In this Gupapuyngu course, which forms part of the study of the Aboriginal Australian in Northeastern Arnhem Land, the starting point has been a concern for the way in which language relates to social structure and culture. A different starting point in no way implies that the aims of traditional foreign language courses are inappropriate. Indeed, the approach taken in this course should complement the in-depth study of a foreign language at secondary school.

It is assumed that the majority of pupils who study the course will speak English as their first language. Through a study of Gupapuyngu, at times compared and contrasted with English, the course sets out to achieve the following aims:

1. To show that Gupapuyngu and other Aboriginal languages are viable and complex means of communication.
2. To demonstrate that the sound systems of languages may differ.
3. To discover that thought forms of languages may differ.
4. To examine the social significance of language.
5. To discover the significance of sign language and other forms of non-verbal behaviour.
6. To appreciate how languages that have never been written may be reduced to writing.
7. To identify some of the universal characteristics of human languages.

Although a recent survey in New South Wales has shown that most parents want their children to learn a foreign language at school, it seems unlikely that many of them would see much sense in learning an Aboriginal language in great detail. One of the major languages of the world, e.g. Spanish, would be a better choice for a prolonged and in-depth study. A great deal of work will therefore need to be done to convince parents that a short alternative language course has much to offer. Any evaluation of the course will need to take into account the perceptions of parents about the worth of the course.

Drop-out rates in foreign language courses in secondary schools throughout Australia have been fairly high. The short course in Gupapuyngu, which is based on 100 words, will not appear as formidable as other foreign language courses but it nevertheless may fail to sustain the interest of students. Teachers and pupils who are involved in the initial trialling of the materials will be asked to record their interest levels in the course as it proceeds.

A major problem during the evaluation of the language module will be to determine the factors which contribute to the acceptance or non-acceptance of the course and the interest or lack of interest generated in the lessons. It will be important to establish whether some of the design features of the course, which were used to achieve certain outcomes, are well received by teachers and pupils. For example, the skills of listening, speaking, reading and writing are introduced simultaneously because the emphasis in this short course is on discovering things about language through each of the four basic skills. In order to make the discoveries, children are introduced to new vocabulary and structures at a controlled rate. There is a balance between deductive and inductive processes in the design of the lessons. On occasions, a rule is stated and pupils are expected to see how it applies. On other occasions, pupils are expected to discover the rule for themselves. These few examples of the design features of the course will suffice to show how difficult it is to evaluate the effectiveness of a language lesson. If new vocabulary cannot be mastered in a lesson, this could affect pupils' capacity to make inferences from the new sentence structures. Then again, if a new sentence structure is introduced through listening activities, as against reading activities, it may be more difficult to grasp. All of the design features of a lesson interact to produce an overall effect. When the course is being trialled, teachers will be expected to reflect on the various design features in the lessons and make some attempt to identify the variables which contribute to the success or otherwise of a lesson.

Formative evaluation of the course by professional linguists will be undertaken. During the preparation of a first draft of the course, Dr. Ken Hale from the Massachusetts Institute of Technology read the materials and provided extremely useful comments on both the content and the approach taken to teaching/learning strategies. For several reasons, transformational exercises have not been stressed as much as they usually are in recently produced foreign language courses. To start with, Gupapuyngu is a highly inflected language in which word order may be varied considerably without affecting meaning. Many of the variations in word order which occur in the course could be termed quasi-transformational exercises. It is only as competent linguists examine the course that the author will know whether children are being exposed to exercises which will give them deep insights into the structure of language.

Miss Beulah Lowe, the linguist who has made a scientific analysis of Gupapuyngu, checked the preliminary draft of the course and made numerous suggestions on how to improve it. In fact, many of the approaches adopted in the course are based upon Miss Lowe's Gupapuyngu grammar. Several Aborigines, including Burrumarra, Badaltja, Ngändama and Djirrimbilpilwuy, have been through the exercises to check them for accuracy.

Questionnaires have been developed for children, parents and teachers to obtain feedback when the materials are trialled. The questionnaire for children, for instance, has six sections:

1. Personal information.
2. My desire to learn a foreign language.
3. What I think about foreign language study.
4. What I think about the use of languages in Australia.
5. What I know about languages.
6. Information about myself.

When constructing the questionnaire, items were generated from texts about the teaching of foreign and second languages. Section 5, which deals with knowledge about languages, examines some of the basic concepts in the Gupapuyngu course. By using experimental and control schools in 1978 for the trialling of the course, the questionnaire for children will be used for pre- and post-tests.

## CONCLUDING REMARKS

In most curriculum development projects, the developers work on the assumption that teachers who will implement a particular project know something about the content of the curriculum. Unfortunately, most teachers who will use the Aboriginal Australian in Northeastern Arnhem Land will know little or nothing about the people being studied. Even if they claim to know something about Aboriginal society and culture, the chances are that this knowledge will not include kinship and language. The authors of the MACOS materials believe that teachers must undergo a course of training to gain some familiarity with the content of the curriculum but more particularly with the teaching strategies. The University of Georgia's anthropology project endeavoured to find out whether a knowledge of anthropology was essential before attempting to teach the curriculum. They found that teachers did not need to have a background in anthropology for their students to learn from the materials, but that conclusion was based on evaluations involving teachers who chose to use the materials. As they point out, many teachers may be reluctant to use the curriculum. The Arnhem Land curriculum materials have been designed in such a way that they do not call for a great deal of background knowledge from the teacher. Where the teacher will exercise his or her skill is in the teaching/learning strategies. When the curriculum is trialled in 1978 and subsequent years, it will be important to determine whether special training is required for teachers who wish to use the curriculum.

## REFERENCES

Barnes, J.A. Inquest on the Murngin. London: Royal Anthropological Institute of Great Britain, 1967.

Notes from here and overseas: Learning Foreign Languages - Popular with English-speaking Parents. Australian Educational Researcher, Vol.3, No.4, 1976, p.22.

Stenhouse, L. et al. The Humanities Project. London: Schools Council Publications, 1970.

Taba, H. et al. A Teacher's Handbook to Elementary Social Studies. Reading, Massachusetts: Addison-Wesley, 1971.

University of Georgia Curriculum Project. Social Education, 1972, p.747.

# EVALUATION OF EARLY CHILDHOOD PROGRAMS

## - WITH SPECIFIC REFERENCE TO THE MT. DRUITT EARLY CHILDHOOD PROJECT[1]

John Braithwaite

Evaluation of educational programs has received heightened attention by educators over the last decade if we consider as evidence the plethora of articles that have been published about evaluation processes and procedures. Yet in spite of the many technical advances that have been made in educational evaluation, there are still many areas that have not received the attention their importance warrants.

One such area concerns the evaluation of early childhood* programs wherein many studies have been reported, but few have presented findings which provide definitive guidance to educators and policy makers. Partly the failure to present findings which are unequivocal, may be attributed to the complexities associated with gathering data on children's behaviours and performances in early childhood settings and partly to the comparative absence of measures** which are both reliable and valid. The combination of factors such as these, together with the marked variability in the behavioural and performance characteristics of four year old children, makes the evaluation of early childhood programs not an easy task.

Yet when comprehensive evaluation studies have been undertaken, many program developers have challenged the inconclusive or negative findings that have been cited. The critics of the evaluation studies of programs such as Head Start and Follow Through, claimed that the evaluators often failed to evaluate the stated objectives of the various programs, did not resolve which dependent and independent variables should be included in the studies, included non-equivalent comparison groups, instituted changes of instrumentation during the data collection periods, used inappropriate monitoring techniques or generally neglected the potential contribution of systematic process data collection.

Salutary as it might be for the developers of these programs to make the above criticisms, there is comprehensive evidence that would also indict the program developers for their sins of omission and commission. Often the program developers failed to frame their objectives carefully, had little knowledge of how the learning experiences that they provided were supposed to affect children's learning and above all, generally failed to give the necessary guidance and support to the evaluators in the design of their evaluation studies.

Fortunately, in devising the evaluation model that was to be implemented in the Mt. Druitt Early Childhood Project, the legacies of the previous evaluation attempts have influenced our approach to the evaluation of the Project. The greater realisation by program developers of the benefits of formative evaluation and of the possible contributions of evaluators to the overall development of the Project, has been reflected in many of our evaluation practices. It is thus our intention to raise certain evaluation issues which have confronted us in the course of the Project's development and evaluation and to show how we have attempted to take account of the dynamic, multiple goals of the Project.***

## THE SETTING OF THE MT. DRUITT PROJECT

The setting of the Project has been described in detail elsewhere.[2] In summary, the Project has attempted to develop, implement and evaluate five different early childhood programs within the Mt. Druitt area of western metropolitan Sydney. This area was chosen for a number of reasons - a) its population has been classified as disadvantaged on a range of different criteria; b) the background variables of the families in the area are relatively homogeneous; c) there are sufficient children in the appropriate age levels to mount such a study; d) the school administrators were willing and supportive of the stated objectives of the Project. The Project was funded for its first phase for a three year period commencing April, 1975. In 1975 and 1976 the training of teachers with the concomitant program developments occupied most of our time. This year 1977, sees data being collected for the evaluation of the Project. Currently there are approximately four hundred and fifty children identified as being members of the evaluation cohort.

---

* "Early childhood" is used here to cover the age ranges from 2 years to 7 years of age.

** "measures" is used here as a generic term embracing the full range of assessment techniques.

*** See Appendix A for Project's goals.

## SOME CONSIDERATIONS CONCERNING THE EVALUATION OF THE PROJECT*

When we first considered the various approaches that we could follow in undertaking the evaluation of our Project, we were conscious of the multiplicity of different models that have been advanced as possible approaches for the evaluation of early childhood programs. The literature on evaluation methodology has identified over thirty different evaluation 'models'[3] each of which purports to advance new approaches to evaluation. As a consensus regarding which elements should be included in an evaluation model of early childhood programs has not yet been reached, it is necessary to outline the criteria which were paramount in our thinking about how we were going to evaluate the Project.

*First*, we believed that we had to adopt a decision-oriented approach as we were committed to providing our various audiences with answers to specific policy questions such as -

> "What particular combinations of resources are most cost-effective for increasing educational growth among our population?"

Answers to questions such as this are important for educational administrators and it was essential that such answers be based on sound empirical evidence.

*Second*, the Project had to evaluate whether each program can attain its stated objectives and to determine the effects each program has on designated sub-groups of children.

*Third*, the evaluation approach had to be flexible enough to monitor the inevitable changes in programs that occurred as ideas and personnel changed. The realization that over a period of time the developments in the Project are dynamic rather than static, was an important influence on our thinking.

*Fourth*, the variables that we chose to collect data on had to reflect Cooley and Lohnes' four measurement domains: 1) initial characteristics of learner, 2) contextual variables, 3) instructional dimensions, and 4) outcomes.[4]

*Fifth*, we approached our evaluation task on a broader conceptual base than is usual in educational evaluations. This was necessary because the interrelationships between the children, their parents, the school and its community have marked effects on children's development at this particular age level. In essence we have adopted what Bronfenbrenner[5] has termed *ecological* evaluation; for we have endeavoured to gather data on the characteristics of the learners and the surroundings in which they live out their lives, and on the interrelationships that exist between the learners and their environments.

*Sixth*, data had to be collected which examines the processes within each program as well as product data in each program.

In stating the above, it is important to note that the criteria listed were not overtly stated before we commenced the Project. Rather like many practices, they evolved during the initial phases of the Project's development.

An additional point must be made concerning our approach to evaluation and that concerns the possible in-built bias that can occur when program developers become evaluators. We have been extremely conscious of this danger and have involved independent personnel who have had no association with the Project before their involvement, to comment upon and evaluate particular aspects of the Project. Certainly one advantage that we have gained through adopting both roles of developer and evaluator has been the ready access to valuable formative data in the Project's development phase.

## EVALUATION GOALS OF THE PROJECT

The above criteria have been articulated into the following evaluation goals:

1. To specify overall, and separately for different kinds of children, the short-term and long-term, intended and unintended effects on pupils, family and school of the five programs.
2. To contribute to the understanding of basic processes in the growth and development of young children. This understanding may be related to educational, medical, psychological and sociological disciplines including understanding better the processes of educational evaluation itself.
3. To describe the processes that occur within each of the five specified programs when functioning as intended.
4. To specify what administrative, fiscal and educational steps have to be taken in order to ensure the effective functioning of the five specified programs.

The design of the study allows for data to be collected at the times specified in Figure 1.

---

* For details of the evaluation model see Ball, S., & Braithwaite, J. "A Working Evaluation Document", ERIC ED 129434.

| Groups | Years | | | | |
|---|---|---|---|---|---|
| | 1975<br>(One term) | 1976 | 1977<br>Major evaluation conort | 1978 | 1979 |
| C1 | Nov (K) | Feb-Nov (Year 1) | | | |
| C2 | | Feb-Nov (K) | Feb-Nov (Year 1) | | |
| C3 | | | Feb-Nov (K) | Feb-Nov (Year 1) | |
| EX1 | | | Feb-June-Nov (PS) | Feb-Nov (K) | Feb-Nov (Year 1) |
| EX2 | | | Feb-June-Nov (PS) | Feb-Nov (K) | Feb-Nov (Year 1) |
| EX3 | | | Feb-June-Nov (PS) | Feb-Nov (K) | Feb-Nov (Year 1) |
| EX4 | | | Feb-June-Nov (PS) | Feb-Nov (K) | Feb-Nov (Year 1) |
| EX5 | | | Feb-June-Nov (PS) | Feb-Nov (K) | Feb-Nov (Year 1) |
| C4 | | | | Feb-Nov (K) | Feb-Nov (Year 1) |

PS, K, Year 1 = Preschool, Kindergarten and Year One
Feb, June, Nov. = Months when major assessments occur
C1 - C3 = Major comparison groups (pre Preschool cohort)
C4 = Major comparison groups (concurrent with preschool cohort)
EX1 - EX5 = Five groups from the five programs

FIGURE 1. Experimental and comparison groups indicating the periods of their major assessments

## EVALUATION ISSUES

As stated above, it is the intention of this paper to raise specific problems that early childhood evaluators have to resolve in most evaluation studies. In practice these problems have centred on *methodological* problems and *measurement* problems. This is not to state that these were the only problems that confronted us in carrying out our evaluation tasks. Rather, the reality of the Project has heightened the interrelations between each problem and such interactions have often created further problems to be solved.

Some of these problems are presented in the following pages as questions which we have had to answer.

## WHAT ARE THE OBJECTIVES OF EARLY CHILDHOOD PROGRAMS IN AUSTRALIA?

Perhaps more so at this level than at any other level of education, stated objectives have been more conspicuous by their absence than by their presence. Terms and catch phrases abound aplenty but when they are analysed for their substantive content and possible application, obfuscation quickly settles in. Teacher educators, directors of preschool centres and teachers in preschools often would state conflicting views concerning the primary objectives of preschool education in Australia. Maybe this lack of clarity about the objectives reflects either a transitional stage in Australian preschool practices or a lack of concern on the part of the educators about the necessity to state objectives. Whatever the cause, this lack of specification causes problems for the evaluator and the statement that each centre "has differing objectives" and "you cannot specify all your objectives" does not help the evaluation process - especially when the organisational/process evidence that has been collected reflects a great similarity in organisation and the learning experiences that are provided in each sessional preschool. Fortunately, there are indications that the preschool educators are realising the value of clarifying their objectives so that evaluations of their programs can be presented in ways that have substantive merit and so will possibly influence the decision makers towards the attainment of the preschool educators' desired goals.

From our perspective, we could only attempt to specify objectives that apply to our programs and thus the generalizability of results can be limited by such a procedure. However, we believe that having loosely stated objectives which perhaps would allow for more generalizability, is less valuable than having more specific objectives. Although generalizability is restricted by choosing the latter alternative, the opportunity for a valid evaluation of our objectives is compensation enough. Certainly by adopting such a procedure we can get closer to answering a major question: "Did we do what we set out to do?"

## SHOULD ALL OUR DATA BE 'HARD' DATA?

When we first began the task of critically reviewing and trialling the potential measures of our objectives, we were concerned about the paucity of measures for children of our targeted age range, which met the traditional canons of reliability and validity. Further, we were concerned at the imbalance in our data collection procedures which would follow by using only established measures which were reliable and valid. Such a procedure would have resulted in completely neglecting data collection in the affective domain. Consequently, we were forced to consider the relative value of collecting 'soft' data which would amplify the more objective data that was to be collected. Essentially questions such as -

> "What is the relative value to an evaluator of a mother stating that she has seen positive changes in a child's behaviour since attending preschool compared with a rating on a scale which is statistically non-significant?"

became matters of debate. We realized that there was a large body of data which could not be classified as objective data, but which nevertheless became essential evidence as to the attainment of the Project's goals.

To resolve this debate we settled on the following as criteria for data collection. First, whether the collection of the data was a feasible task from the standpoints of practicability, costs and personal impact on the potential data sources. Second, whether the potential data can be subjected to appropriate analytical procedures. Third, whether we were going to have a collection of data whose complexity precludes us from ever providing appropriate empirical conclusions within our funding period. Fourth, we have had to be continuously aware of the community's legitimate concern over becoming the subject of continuous enquiry by 'those people from the university'.

## HOW RELIABLE AND VALID ARE THE MEASURES THAT ARE USED?

Any evaluator who has attempted to gather reliable and valid evidence from preschool aged children knows too well the number of times that reviews of specific measures that show reliability and validity are "not known" or are "highly questionable". It is all too apparent that obtaining reliable measures from preschool aged children who have never been in a test taking situation before is difficult. Hence we have had to question the administration procedures of the measures that we have employed, the validity of the measures and their potential reliability. Unfortunately, we must state that many of the measures we have trialled have had to be discarded on one or more of the above deficiencies.

To overcome these problems, we have embarked upon two courses. First, we have developed with the teachers, measures which are criterion referenced and these measures form the basis of our program measures. Second, we consciously avoided in the initial phase of the Project having to devise specific Project measures as this constituted yet further re-invention of the wheel. Unfortunately, after we had evaluated most standardized measures for children of this age level, we reluctantly concluded that we had to devise some measures to evaluate our objectives. This development occupied us for a great deal of our time during 1976 and hopefully some of the measures will have greater replicability in future studies of a similar nature in Australia. Thus the answer to our search for reliable and valid measures has been to take a course that we hoped we would not have to embark upon, but the necessity to have measures to answer our specific Project questions overruled our hesitancy about developing new measures.

## HOW DO WE RECONCILE THE DIFFERING DEMANDS MADE BY VARIOUS AUDIENCES?

Each of the various audiences for whom the Project has to present results has asked for differing products at the completion of the Project. Policy makers want answers to questions that are concerned with cost-effectiveness, program effectiveness and possible educational benefits. Educationists in other settings want answers to questions that are related to trait-treatment interactions, intended and unintended outcomes, implementation and evaluation procedures and problems. Parents want to know if it is worthwhile to their child to attend a sessional preschool and how these benefits might show up. The funding body has its particular needs in terms of policy input documents. Additionally, the Project staff want answers to questions that have come to interest them, such as the perceived value of personal development courses for parents and high school students. All of these constituents have a legitimate right to have their questions answered. Thus, as evaluators, we must attempt to identify these audiences and their potential requests, which they often have not expressly stated, and to present the findings in appropriate ways. Moreover, as it is more than likely previously unidentified demands will appear, the evaluation model must be flexible enough to respond to these emerging demands.

Another problem which inevitably occurs in the very best evaluation models is what to do when a carefully considered element disintegrates before your eyes. For example, we have had to cope with teachers whom we have specially trained transferring and so on. How have their actions affected our results? These and other issues necessitate an evaluation model which can maintain its essential integrity but react positively to spontaneous pressures such as those noted. Keeping an even keel in such a tacking situation requires the evaluator to keep the finish line always in sight so that the demands of the various audiences can be met.

## HOW DO YOU ALLOW FOR THE GREAT VARIABILITY IN CHILDREN'S PERFORMANCE AND BEHAVIOURS AT THIS AGE LEVEL?

We know how daily specific children's behaviours and value judgements at this age level can be. For example, our initial investigations into locus of control measures gave us a variety of responses, when we asked the same questions in a variety of formats. Which response was the reliable and valid one became our problem. Similar results were obtained with other measures.

A further complication is caused by developmental variabilities in children's performance on standardized tests. For example, a test that measures visual perceptual abilities may be developmentally appropriate at age five but not at age four. How to resolve such an issue? If we devised the test to include items measuring developmentally appropriate tasks at each age level we ran the risk of establishing test fatigue. By splitting the test into two parts we encountered the further variability in responses caused by specific personal events in the child's life on the first or second day of the test's administration.

These are but some of the issues that we have had to resolve in order to ensure that our measures were both reliable and valid.

## SOME OTHER PROBLEMS

Attrition within our Project's population has already caused us problems. Such problems will be magnified if we continue on with our longitudinal follow-up of the 1977 cohort. Ways of ensuring that an accurate account is kept of the possible changes of address of the children and their parents have had to receive close attention.

We have already encountered the legitimate demands of the teachers and administrators for results from our data collection procedures. Certainly when we have presented them with the evidence that we have collected to date, we have sensed a greater involvement on their part towards the Project. Yet the inevitable delay between data collection, analysis and reporting means that we cannot always present findings as quickly as we would wish. Again at times there are certain data that we do not wish to present in order to control for possible contamination. Balancing the ethical demands of reporting data wherever possible and not influencing teachers'/administrators' behaviours is not an easy task. We have a commitment to providing information back to the schools and the community, otherwise we are neglecting our responsibilities to them. We must leave something of value behind so that the society can appreciate the potential benefits from field research.

## CONCLUSION

The above questions are but some of our concerns with developing a comprehensive evaluation model of early childhood education. At the back of our minds is always the question - "Does the sum of our evaluation activities always equal the whole child, the whole program and the whole project?" We are inclined to think not, as while our model has attempted to account for the quintessence of our Project, the very nature of the possible interactions and the ways in which external forces can affect our developments, suggest that we might never achieve all of our evaluation goals. This is one of the prices we have to pay for evaluating educational programs in the public school system and its communities. However, we believe that the evidence that we shall eventually present will have greater impact since it will reflect the uncertainties and conflicting demands imposed upon all within a mass education system.

Undoubtedly we must be reconciled that the pluperfect subjective is unfortunately and notoriously unresearchable.

John Braithwaite
Field Director
Mt. Druitt Early Childhood Project
School of Education
Macquarie University
North Ryde 2113

## REFERENCES

1. This paper reflects some of the evaluation concerns of the Project staff, hence many of their individual views are reflected within it. However, the author accepts responsibility for the ideas that are presented.

2. Braithwaite, J. (1977) "The Mount Druitt Early Childhood Project", Education News, 16, 1, 18-21.

3. Antonoplos, D. P. (1977) Evaluation Reconsidered: Pied Pipers of Pedagogy. Paper presented at AERA Meeting, April, 1977.

4. Cooley, W., & Lohnes, P. (1976) Evaluation Research in Education - Theory, Principles and Practice, New York: Irvington Publishers, p.5.

5. Bronfenbrenner, U. (1976) "The Experimental Ecology of Education", Educational Researcher, 5, 9, (October 1976), 5-15.

## APPENDIX A

## GOALS OF MT. DRUITT EARLY CHILDHOOD PROJECT

- Provide an environment which will help the child to develop his/her physical and intellectual abilities through a widening range of experiences.
- Give each child feelings of enjoyment and success in a schooling environment which should develop his/her self identity, initiative and confidence.
- Involve the community and parents in the education of their children by helping them to see themselves as partners in the educative process.
- Develop the child's conceptual and language abilities.
- Encourage the child to develop appropriate social behaviours.
- Develop the unique creative potential of each child by encouraging an active curiosity about the world in which s/he lives.
- Provide support to other agencies in caring for the child's health, nutritional and other needs.
- Evaluate the differential outcomes of the five programs to preschool education.
- Provide the guidance necessary for the continuity of approaches through the upper grade levels.

# ANXIETY, ACHIEVEMENT

# AND

# ANTIPATHY

RUSSELL A. DOCKING
Murdoch University

This paper outlines the background to a research study that will seek to determine whether the antecedent of ineffective communication - cognitive incongruence or antipathy, is related to raised state of anxiety, and will replicate an earlier study of antipathy and achievement. A response paradigm derived from Rogers self theory is considered as a possible explanation of the frequently observed negative correlation between anxiety and achievement.

In 1951, Rogers laid a structural foundation for the phenomenological study of personality and behaviour in nineteen propositions. These statements tell us much about normal and abnormal behaviour, and have significant implications for research strategies, stressing as they do the need for interactional studies and self-perception measures. They have much to say about the meaning of "reality" and the way in which one responds to the experience of reality, and hence have implications for communication theory and learning theory.

In the explanation accompanying Proposition Sixteen, Rogers describes a mechanism (attributed to Hogan), of dealing with incongruity between an experience and the self, (Rogers, 1951, p.516). This mechanism is shown in diagramatic form in Figure 2.

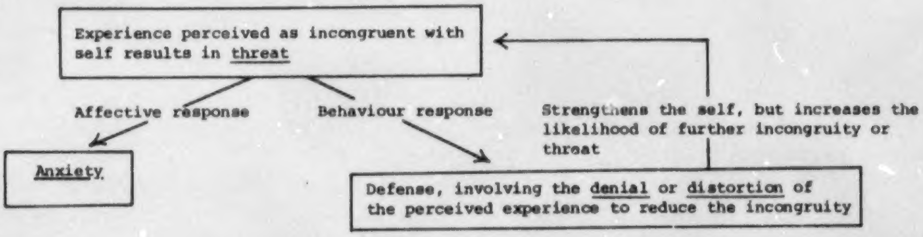

Experience perceived as incongruent with self results in threat
Affective response
Behaviour response
Strengthens the self, but increases the likelihood of further incongruity or threat
Anxiety
Defense, involving the denial or distortion of the perceived experience to reduce the incongruity

Figure 1. MECHANISM OF RESPONSE TO INCONGRUENCE (AFTER ROGERS, 1951)

Spielberger has a similar model linking anxiety, defensiveness and behaviour (Spielberger, 1966a, p.17) that is a little less useful as it does not use threat as an intervening variable. The cognitive appraisal stage is highlighted by Spielberger, however, making it clear that cognitive style plays a significant role in threat perception and hence anxiety. (Lazarus and Opton, 1966, pp.229-230).

As perception-self incongruity can arise in a communication situation (Barnlund, 1968), it follows from Figure 1 that if the stimuli of the teacher are incongruent with the expectations of the student, there will be two observable outcomes:

1) Affective: the raising of anxiety
2) Behaviour: the distortion or denial of learning.

The second of these outcomes - the effect of staff-student incongruity upon student achievement - has been studied by the author and reported elsewhere (Docking, 1976a, 1976b). The first outcome - raised anxiety, is the particular concern of this study, which will attempt to clarify the following issues:

1) the extent to which staff-student intellectual incongruity effects anxiety

and 2) the extent to which the response mechanism described in Figure 1 accounts for the observed negative correlation between anxiety and achievement.

This study involves measures of staff and student cognitive styles from which indices of antipathy are determined, student anxiety, and student achievement - the last of these enabling a replication of the previous research on antipathy and achievement.

## Anxiety and Antipathy

What research evidence is there that anxiety is related to incongruity? Given that incongruity of cognitive structure is revealed through verbal and non verbal communication between one person and another, (De Vito, 1970, p.86; Fulcher and Anderson, 1974, p.16), efficacy of communication can be used as an indication of antipathy. There are a number of references in the literature relating difficulty in communication to raised anxiety levels (Bettinghaus, 1973; Duncan, 1968; Griffith and Patten, 1971; Hansen, 1974; Horenstein and Gilbert, 1976; Ragsdale, 1976; Smith, 1976), but none are so explicit as Grinker.

> ... the best way to produce anxiety was to impede communication ... methods of producing anxiety, such as threatening harm, or shaming, were relatively ineffective. (Grinker, 1971, p.136).

There is also a body of indirect research evidence that anxiety is related to difficulty in communication and hence to antipathy. This is found in studies of behavioural correlates of anxiety (either teacher or student behaviour). Notably, one finds anxiety reduced in feedback-rich environments (Gaudry and Bradshaw, 1975; Hoge and Luce, 1976; Levy and Kellmer-Pringle, 1969; Lewis and Adank, 1975), or where communication is encouraged (McKeachie, Pollie and Spiesman, 1955), or with teachers who exhibit communicative, supportive behaviours (Cameron, 1975, Zimmermann, (1970) Other studies also suggest the anxiety-communication relationship (Boak and Conklin, 1975; Doyal and Forsyth, 1973; Heys, 1974; Joesting and Whitehead, 1976; McKeachie, 1961; Miller and Michael, 1972; Rassaby and Winkler, 1976; Sinclair, Heys and Kemmis, 1974).

These direct and indirect references to anxiety and communication, summarised in Table 1, lend substantial support to the hypothesis that cognitive incongruity (antipathy) is related to anxiety.

Table 1. REVIEW OF RECENT STUDIES LINKING ANXIETY AND ANTIPATHY

| | | |
|---|---|---|
| Bettinghaus, 1973 | | Communication serves to reduce anxiety. |
| Duncan, 1968 | | Anxiety related to "concern how to communicate". |
| Griffith and Patten, 1971 | | When communication is blocked, anxiety is generated. |
| Hansen, 1974 | STAI | A-State reduced significantly through feedback in computer exercise. |
| Horenstein and Gilbert, 1976 | AAC | Self disclosure raises anxiety in others, but not self. (revelation of incongruence) |
| Ragsdale, 1976 | Welsh | High correlation between anxiety and hesitation in normal communication. |
| Smith, 1976 | MAS | Anxiety raised by "double blind" communication. |
| Grinker, 1971 | | Blocking communication raises anxiety levels. |
| Gaudry and Bradshaw, 1975 | TASC | Anxiety lower in examination procedures that enhance communication. |
| Hoge and Luce, 1976 | TASC | Communication of expectations raises anxiety. |
| Levy and Keller-Pringle, 1969 | GASC,<br>TASC | Anxiety levels higher in a traditional education form vs child-centred form. |
| Levy and Adank, 1975 | GASC,<br>TASC | Anxiety lower in individualized instruction than traditional instruction. |
| McKeachie, Pollie and Spiesmann, 1975 | TAS | Opportunity to communicate opinions on exam reduced anxiety. |
| Cameron, 1975 | TASC | Lowered anxiety with "warm, nurturant, challenging and non-dominating teachers". |
| Zimmerman, 1970 | SAQ | Negative correlation between anxiety and teacher positive reinforcement and teacher directiveness. |
| Boak and Conklin, 1975 | HSPQ | Interpersonal skills slightly related to anxiety. |
| Doyal and Forsyth, 1973 | TASC,<br>MAS | Anxiety of teacher correlated with anxiety of children. (Mutual antipathy) |
| Heys, 1974 | TASC,<br>Others | Anxiety component "Lack of confidence in meeting the expectations of others". |
| Joesting and Whitehead, 1976 | STAI | High correlation between A-State and A-Trait and Confusion-Bewilderment Mood State. |
| McKeachie, 1961 | | Student anxiety related to similarity of teacher cues to student expectations. |
| Miller and Michael, 1972 | SAMS | Anxiety highly correlated with alienation (feeling of isolation and rejection). |

Table 1 (Contd)

| | | |
|---|---|---|
| Rassaby and Winkler, 1976 | GASC | Hypothesised anxiety to be raised in unstructured classes (not confirmed). |
| Sinclair, Heys and Kemmis, 1974 | | Anxiety related to reduced accessibility to cues. |

Of course, these studies do not directly link anxiety and antipathy, and this will be the primary aim of this research.

## Antipathy and Achievement

The behaviour component of Rogers mechanism of response to incongruity suggests that achievement will be negatively correlated with antipathy. This, the major hypothesis of the thesis by the author, (Docking, 1976b), was supported in an extensive study of students undertaking a University Physics programme.

Support in other writings is mainly in the form of indirect evidence, (Baird, 1971; Boak and Conklin, 1975; Centra and Rock, 1971; Cohen and Berger, 1970; Hansen, 1974; McKeachie, 1961; Pervin, 1967; Pervin and Rubin, 1967). Lavin (1965) in reviewing research on achievement at tertiary level concluded:

> ... the degree of congruence in student-teacher values, attitudes and expectations is directly related to the academic performance of the student. Furthermore, this generalization holds even when the student-teacher similarity involves what seems to be task-irrelevant criteria. (p.144)

These references, summarized in Table 2, lend support to the hypothesis that antipathy is related to achievement.

Table 2. REVIEW OF RECENT STUDIES LINKING ANTIPATHY AND ACHIEVEMENT

| | |
|---|---|
| Docking, 1976b | Similarity of cognitive style related to achievement. |
| Baird, 1971 | Observed need for research on interaction between student characteristics and teaching style. |
| Boak and Conklin, 1975 | Teacher interpersonal skills significantly related to student achievement. |
| Centra and Rock, 1971 | Achievement related to quality of instruction. |
| Cohen and Berger, 1970 | Student-faculty interaction significantly related to achievement. |
| Hansen, 1974 | Feedback in computer exercise increased performance in high reasoning students. |
| McKeachie, 1961 | Observed need for research on interaction between student characteristics and teaching. |
| Pervin and Rubin, 1967 | Self-college similarity related to "drop out" rates. |
| Lavin, 1965 | Congruence between teacher and student values related achievement. |
| Tudor and Holmes, 1973 | Distorted recall did not accompany high anxiety. (However, the measure of anxiety was taken at the time of learning, not recall). |

These studies link antipathy with achievement, and this will be replicated in the course of this study.

## Anxiety and Achievement

There is no lack of research literature dealing with anxiety and achievement (Gaudry and Spielberger, 1971; Spielberger, 1966b). Recent studies are described in Table 3.

Table 3. REVIEW OF RECENT STUDIES LINKING ANXIETY AND ACHIEVEMENT

| | | |
|---|---|---|
| Duffey and Martin | STAI | Primary school. High A children scored better than Low A children. |
| Dubey, 1976 | Sinha | Tertiary students. High A students perform better than Low A students. |
| Geisen, Ehlers, Heyer, 1976 | MAS | Tertiary students. Performance correlated negatively with achievement. |
| Giffore, 1977 | Albert-Haber | Teachers in-service. r-.23 ( 0.02). |
| Hansen, 1971 | STAI | Tertiary students. High A students have higher error rate. |
| Heinrich, Stephenson and Spielberger, 1976 | STAI | Graduates. r-0.03 to -0.33. A-State and achievement. |
| Heys, 1974 | TASC | Primary school. r-0.09 to -.41 (girls) |
| Hoge and Luce, 1976 | TASC | Primary school. r-.31 ( 0.01) |
| Joesting and Whitehead, 1977 | STAI | Tertiary students 4-.26 to -.24 A-State and performance, r-.2- to -.15 for A-Trait. |
| Kanenka, Nedakantar and D'Souza, 1976 | MAS | High School, 4-.11. |

Table 3 (Contd)

| | | |
|---|---|---|
| Lewis and Adank, 1975 | GASC & TASC | Primary school, r-.16 (ns) individual instruction, r-.36 ( 0.001) group instruction. |
| Müller and Michael, 1972 | SAMS | Tertiary students, r-.13 to -.32. |
| Vagt and Kühn, 1976 | MAS, ATS | Middle school. High A students prepare more, but perform no better than Low A students. |
| Zimmermann, 1970 | SAQ | Primary school, r-.24 to -.45. |

There is a general agreement that anxiety is negatively correlated with achievement, with Dubey and Martin dissenting. It is widely accepted that the relationship between anxiety and achievement is not a simple linear one, but rather it is likely to be an inverted U shape. (Malmo, 1966). At low levels of anxiety, increasing anxiety increases performance (as anxiety acts as a drive), with a levelling or saturation effect at optimum performance, followed by a decline in performance for higher levels of anxiety. The mechanism of this decline is not clear, perhaps it is cue interference (Hebb, 1955), or alternatively, perhaps the increasing anxiety reflects an increasing uncertainty that also results in lowered performance (See figure 2).

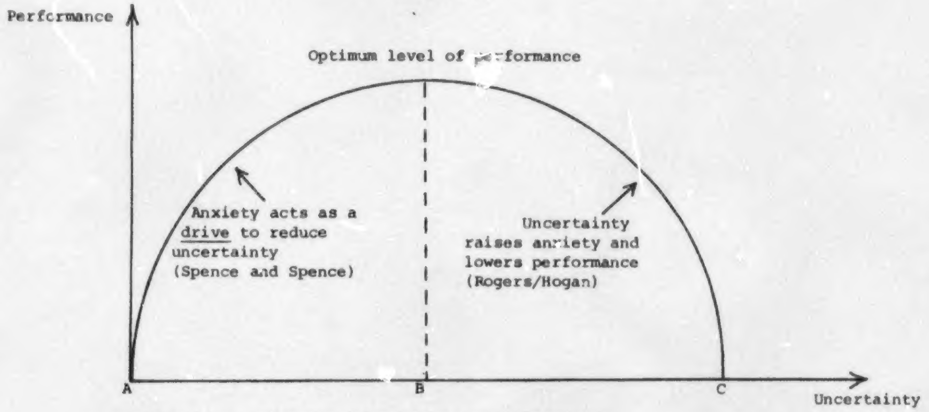

Performance
Optimum level of performance
Anxiety acts as a drive to reduce uncertainty (Spence and Spence)
Uncertainty raises anxiety and lowers performance (Rogers/Hogan)
A
B
C
Uncertainty

Figure 2. ANXIETY, UNCERTAINTY AND PERFORMANCE

This explains the contradictory results in Table 3. In region AB, achievement and anxiety are positively correlated as anxiety acts as a drive; in region BC, anxiety accompanies a defensive behaviour that distorts and lowers performance, and is therefore negatively correlated with performance. The results of research relating anxiety and achievement will therefore depend on the drive vs defense response of the group being tested.

The question of causality between anxiety and achievement is put aside by the alternative hypothesis that both have a common antecedent - ambiguity. This will be the third area of research in the proposed study.

### Antipathy Measures

Antipathy will be based on staff and student responses to the Student Role Orientation (SRO) Questionnaire, developed, validated and utilized in the 1976 study. Antipathy is determined from these using Cattell's $r_p$ coefficient of similarity (Cattell, 1966). These indices will then be reduced to ranks of similarity of each staff member with each student for later analysis. The SRO scales have already been found to be valuable indicators of incongruence related to achievement.

### Achievement Measures

Achievement will be the students standard score grades awarded by each lecturer. This will also be reduced to a set of ranks for each student for further analysis. The use of grades awarded is deliberate as this mark reflects the outcome of the final communication act of the student and teacher and is therefore a valid measure for this study. In addition it is these real measures of achievement that are the basis of student progress.

### Anxiety Measures

Anxiety will be measured by a modified Zuckerman Affect Adjective Checklist (AAC) (Zuckerman, 1960). The selection of this rather obscure measure is based on an extensive review of anxiety measures, that will now be outlined.

The nature of the experiment requires a measure of A-State for each course taken in a semester. The most widely used measure of A-State, Spielberger's STAI, is not suitable as it is a little too lengthy, is insufficiently disguised for repeated administration to sophisticated respondents, and is too restrictive in focus. This study demands the definition of another kind of anxiety somewhere between A-State and A-Trait - general time but specific situation. This is contrasted with the traditional forms defined by Spielberger (1966a) in Figure 3.

| | | Time | |
|---|---|---|---|
| | | General | Specific |
| Situation or Activity | General | A-Trait | A-Periodic |
| | Specific | A-Situational | A-State |

Figure 3. DERIVATION OF ALTERNATIVE FORMS OF ANXIETY

A-Periodic refers to anxiety which is time dependent but situation independent, for example, biorhythms, and other biological clocks.

A-Situational refers to anxiety which is situation/activity dependent but time independent, such as anxiety related to heights or confined spaces, or, as in this study, related to taking a particular course with a particular lecturer. This concept of anxiety has also been proposed by Lamb (1976).

The Zuckerman Affect Adjective Checklist (AAC) is quicker to complete, easier to adapt to measure any of the above forms of anxiety, and is less threatening or obvious in form than the STAI. The twenty one key words of the AAC were embedded in thirty nine adjectives that could be used to describe any tertiary course, and listed in alphabetical order. The new instrument was validated with the STAI (see Table 4) and found to meet the required standards.

Table 4. SUMMARY OF PROPERTIES OF THE MODIFIED AAC

| AAC, STAI Correlations | N = 35 | N = 15 |
|---|---|---|
| A-Trait | .74** | .63** |
| S-State | .70** | .62** |

Other indications of validity/reliability

1. A-situation ranks in courses determined from AAC compared with rank by direct self-report.
   Page's L = 618** m = 23, n = 4.
2. Test - retest reliability over one week N = 34. r = .83**
3. Other correlations

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. A-Trait AAC | - | .52** | .76** | .47** | .74** | .46** |
| 2. A-State AAC | 34 | - | .14 | .62** | .31 | .70** |
| 3. A-Situation - Murdoch AAC | 20 | 20 | - | .53* | .65** | .32 |
| 4. A-Periodic - Murdoch AAC | 34 | 34 | 20 | - | .37 | .45** |
| 5. A-Trait STAI | 32 | 21 | 20 | 21 | - | .61** |
| 6. A-State STAI | 31 | 35 | 19 | 33 | 20 | - |

These correlations fit expectation (i.e. degree of expected similarity).

$^{*}p < 0.05$ $^{**}p < 0.01$

The reliability figure obtained for the modified AAC compares very favourably with those obtained for the STAI (Barker, Wadsworth and Wilson, 1976; Joesting, 1977; Metzger, 1976; Wadsworth, Barker and Barker, 1976) and for the original AAC (Zuckerman, 1960).

The validation of the AAC against the STAI begs the question of the validity of the latter. This is itself an area of controversy, with some reports showing poor correlation with somatic and projective measures (Blankstein, 1976; Salter, Meunier and Triplett, 1976) others showing little distinction between Trait and State measures (Barker, Wadsworth and Wilson, 1976; Endler, Magnusson, Ekehammer and Okada, 1976; Wadsworth, Barker and Barker, 1976). The results of these studies do not go as far as to invalidate Spielberger's measures, nor do they challenge his original concept of the Trait and State forms of anxiety - a distinction that has also been identified by Cattell, Shrader and Burton (1974) through a factorial analysis of the HSPQ. The results reported in Table 4 also support the distinction between A-State and A-Trait for both the STAI and AAC.

The AAC will be used to measure A-situational in each course taken by each student, and these results will be reduced to a set of ranks for each student for further analysis.

## Analysis of Results

Page's L test, a test of linear ranks, will be used to compare antipathy/achievement, antipathy/anxiety and anxiety/achievement ranks (Page, 1963).

## Conclusion

The completion of this study now only requires the administration of the SRO and AAC (one per course taken) to a group of students, the SRO to their teachers, and the collection of semester results. Delays have occurred in the selection of and access to samples that meet the experimental requirements. It is hoped to have preliminary results available at the time of presentation of this paper.

## References

Baird L.L., Teaching styles : An explanatory study of dimensions and effect. E.T.S. Research Bulletin, 1971.

Barker H.R., Wadsworth A.P. and Wilson W., Factor structure of the State-Trait Anxiety Inventory in a non-stressful situation. Journal of Clinical Psychology, 1976, Vol.32, pp.595-598.

Barnlund D.C., Interpersonal Communication : Survey and Studies. Houghton-Mifflin, Boston, 1968.

Bettinghaus E.P., Persuasive Communication. 2nd Ed., Holt, Reinhart and Winston, 1973.

Blankstein K.R., Relationships between Spielberger Trait Anxiety and Lykken Social and Physical Trait Anxiety. Journal of Clinical Psychology, 1976, Vol.32, pp.781-782.

Boak R.T.R. and Conklin R.C., The effect of Teachers' levels of interpersonal skills on Junior High School students' achievement and anxiety. American Educational Research Journal, 1975, Vol.12, pp.537-549.

Cameron B.E., Some relationships between perception of teacher and certain motivational dispositions in Primary Grade pupils. Australian Journal of Education, 1975, Vol.19, pp.287-288.

Cattell R.B., Coulter M.A. and Tsvjioka B., The taxonometric recognition of types and functional emergents. Cattell R.B. Handbook of Multivariate Experimental Psychology, 1966, pp.288-329.

Cattell R.B., Shrader R.R. and Barton K., The definition and measurement of anxiety as a trait and a state in the 12 to 17 year range. British Journal of Social and Clinical Psychology, 1974, Vol.13, pp.173-182.

Centra J.A. and Rock D., College environments and student academic achievement. American Educational Research Journal, 1971, Vol.8, pp.623-634.

Cohen S.H. and Berger W.G., Dimensions of students' ratings of college instructors underlying subsequent achievement on course examinations. Proceedings of the 78th Annual Convention, A.P.A.,1970.

De Vito J., The Psychology of Speech and Language, Random House, 1970.

Docking R.A., Staff-Student Interaction at Tertiary Level. Unpublished Ph.D Thesis. Monash University, 1976a.

Docking R.A., Intellectual Environment and Academic Achievement at Tertiary level. Australian Journal of Education, 1976b, Vol.20 (Paper presented at 1975 AARE Conference, Adelaide).

Doyal G.T., and Forsyth R.A., The relationship between teacher and student anxiety levels. Psychology in the Schools.

Dacey R.S., Manifest anxiety and educational performance. Psychological Studies, 1976, Vol.21, pp.52-58.

Duffey J.B. and Martin R.P., The effects of direct and indirect teacher influance and student trait anxiety on the immediate recall of academic material. Psychology in the Schools.

Duncan H.D., Symbols in Society, O.U.P., 1968.

Endler N.S., Magnusson D., Ekenammer B., and Okada M., The Multidimensionality of state and trait anxiety. Scandinavian Journal of Psychology, 1976, Vol.17, pp.81-96.

Fulcher D.G. and Anderson W.T., Interpersonal dissimilarity and teaching effectiveness : A relational analysis. Journal Of Educational Research, 1974, Vol.68, pp.19-25.

Gaudry E.and Bradshaw G.D., The differential effect of anxiety on performance in progressive and terminal school examinations. Australian Journal of Psychology, 1970, Vol22, pp.1-4.

Gaudry E.and Spielberger C.D., Anxiety and Educational Achievement. Wiley. Sydney. 1971.

Giesen H., Ehlers T. and Geyer M., Angstlichkeit und Leistung: Eine Untersuchung zur Vorhersage der Vorhersagbarkeit. *Zeitschrift fur experimentelle und angewandte Psychologie*, 1976, Vol.23, pp.175-189.

Giffen K and Patton B.R., *Basic Readings in Interpersonal Communication*. Harper and Row, New York, 1971.

Griffore R.J., Correlation between performance in final examination and debilitating anxiety among professional educators. *Psychological Reports*, 1977, Vol.40, pp.954.

Grinker R.R., The psychosomatic aspects of anxiety. In C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966, pp.129-141.

Hagtvet K.A., Worry and emotionality components of test anxiety in different sex and age groups of elementary school children. *Psychological Reports*, 1976, Vol.39, pp1327-1334.

Hansen J.B., Effect of feedback, learner control and cognitive abilities on State anxiety and performance in a computer assisted instruction task. *Journal of Educational Psychology*, 1974, Vol.66, pp.247-254.

Hebb D.O., Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 1955, Vol.62, pp.243-254.

Heys T.A., Anxiety and classroom performance. *Australian Journal of Education*, 1974, Vol.18, pp.184-195.

Hoge R.D. and Luce S., Behavioural and performance correlates of test anxiety in children. *Alberta Journal of Educational Research*, 1976, Vol,22, pp.196-204.

Holroyd K.A., Cognition and desensitization in the group treatment of test anxiety. *Journal of Consulting and Clinical Psychology*, 1976, Vol.44, pp.991-1001.

Horenstein D. and Gilbert S.J., Anxiety, likeability and avoidance as responses to self-disclosing communication. *Small Group Behaviour*, 1976, Vol.7, pp.423-432.

Janiesse M.P. and Palys T.S., Frequency and intensity of anxiety in university students. *Journal of Personality Assessment*, 1976, Vol.40, pp.502-515.

Joesting J., Test-retest reliabilities of state-trait anxiety inventory in an academic setting. *Psychological Reports*, 1975, Vol.37, p.270.

Joesting J. and Whitehead G.I., Personality correlates of two administrations of state-trait anxiety inventory. *Psychological Reports*, 1976, Vol.38, pp.585-586.

Joesting J. and Whitehead G.I. Relationship of state and trait anxiety to grades in educational psychology. *Psychological Reports*, 1977, Vol.40, pp.705-706.

Kanekar S., Neelakantan P. and D'Souza., Anxiety, intelligence and academic performance. *Psychological Reports*, 1976, Vol.38, p.938.

King F.J., Heinrich D.L., Stepherson R.S. and Spielberger G.D., An investigation of the causal influence of Trait and State anxiety on academic achievement. *Journal of Educational Psychology*, 1976, Vol.68, pp.330-334.

Kjerulff K.and Wiggins N.H., Graduate student styles for coping with stressful situations. *Journal of Educational Psychology*, 1976, Vol.68, pp.247-254.

Lamb D.H., Usefulness of situation-specific trait and state measures of anxiety. *Psychological Reports*, 1976, Vol.38, pp.188-190.

Lavin D.E., *The Prediction of Academic Performance*. Russell Sage Foundation, N.Y., 1965.

Lazarus R.S. and Opton E.M., The study of psychological stress: A summary of theoretical formulations and experimental findings in C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966, pp.225-262.

Levy S.G. and Kellmer-Pringle M.L., A longitudinal study of the relationship between anxiety and streaming in a progressive and traditional junior school. *British Journal of Educational Psychology*, 1969, Vol.39, pp.166-173.

Lewis J and Adank R., Intercorrelations among measures of intelligence, achievement, self-esteem and anxiety in two groups of elementary school pupils exposed to two different models of instruction. *Educational and Psychological Measurement*, 1975, Vol.35, pp.449-501.

Malmo R.O., Studies of anxiety: Some clinical origins of the activation concept. In C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966, pp.157-178.

McKeachie W.J., Motivation, teaching methods and college learning. In M.R. Jones (Ed) *Nebraska Symposium in Motivation*, 1961.

McKeachie H.J., Pollie D. and Spiesman J., Relieving anxiety in classroom examinations. *Journal of Abnormal and Social Psychology*, 1955, Vol.50, pp.93-98.

Metzger R.L., A reliability and validity study of the State-Trait Anxiety Inventory. *Journal of Clinical Psychology*, 1976, Vol.32, pp 276-278.

Michael W.B., Michael J.J., Zimmerman W.S. and Miller D.C., *The Development and Validation of an Instrument to Measure Study Attitudes and Methods*. A.P.S. Congress, Canberra, 1972.

Miller D.C. and Michael W.B., *The Relationship of each of six scales of the Study Attitudes and Methods Survey (SAMS) to each of two criteria of academic achievements in community college*. A.P.S. Congress, Canberra, 1972.

Nicholls J.G., When a scale measures more than its name denotes: The case of the Test Anxiety Scale for Children. *Journal of Consulting and Clinical Psychology*, 1976, Vol.44, pp.976-965.

Page E.B., Ordered hypotheses for multiple treatments: A significance test for linear ranks. *American Statistical Association Journal*, 1963, Vol.58, pp.216-231.

Parish T.S., Buntman A.D. and Buntman S.R., Effect of counterconditioning on test anxiety as indicated by digit span performance. *Journal of Educational Psychology*, 1976, Vol.68.

Pervin L.A., Performance and satisfaction as a function of individual environment fit. *Psychological Bulletin*, 1968, Vol.69, pp.56-58.

Pervin L.A. and Rubin D.B., Student dissatisfaction with college and the college dropout: A traditional approach. *Journal of Social Psychology*, 1967, Vol.72, pp.285-295.

Ragsdale J.D., Relationships between hesitation phenomena, anxiety and self-control in a normal communication situation.

Rassaby E.R. and Winkler R.C., Children's anxiety, I.Q. and sex as predictors of teachers' ratings of adjustment to open classrooms. *Australian Educational Researcher*, 1976, Vol.3, No.2, pp.16-20.

Rogers C.R., *Client centred Therapy*, Houghton-Mifflin, 1951.

Salter C.A. and Meunier J.L. and Triplett N.M., Multiple measurement of anxiety and its effect on complex verbal learning. *Psychological Reports*, 1976, Vol.38, pp.691-694.

Sieber J.E. and Crockenberg S.B., The teacher and the anxious child. *Today's Education*, Oct, 1970, pp.76-77.

Sinclair K.E., Heys T.A., Kemmis D.deC., Anxiety and cognitive process in problem solving. *Australian Journal of Education*, 1974, Vol.18, pp.239-255.

Smith E.K., Effect of the double-blind communication on the anxiety level of normals. *Journal of Abnormal Psychology*, 1976, Vol.85, pp.356-363.

Spence J.T. and Spence K.W., The motivational components of manifest anxiety: Drive and drive stimuli. In C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966, pp.291-324.

Spielberger C.D., Theory and research in anxiety. In C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966a, pp.3-20.

Spielberger C.D., The effects of anxiety on complex learning and academic achievement. In C.D. Spielberger (Ed) *Anxiety and Behavior*, Academic Press, 1966b, pp.361-398.

Tudor R.G. and Holmes D.S., Differential recall of success and failure: Its relationship to defensiveness, achievement motivation and anxiety. *Journal of Research in Personality*, 1973, Vol.7, pp.208-224.

Vagt G. and Kuhn B., Zum Zusammenhang zwischen Angstlichkeit und Schulleistung die Berucksichtigung des Ausmasses der hauslichen Vorbereitung auf schulische Prufungssituationen. *Zeitschrift fur experimentelle und angewandte Psychologie*, 1976, Vol.23, pp.163-173.

Wadsworth A.P., Barker H.R. and Barker B.M., Factor scructure of the State-Trait Anxiety Inventory under conditions of variable stress. *Journal of Clinical Psychology*, 1976, Vol.32, pp.576-579.

Zimmerman B.J., The relationship between teacher classroom behaviour and student school anxiety levels. *Psychology in the Schools*, 1970, Vol.7, pp.89-93.

Zuckerman M., The development of an Affect Adjective Checklist for the measurement of anxiety. *Journal of Consulting Psychology*, 1960, pp.457-462.